

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 02:04:13 ; Search time 2348 Seconds
(without alignments)
10151.277 Million cell updates/sec

Title: US-09-673-918a-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgtccaataatccagattct 819

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pi.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vl.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	806.4	98.5	831	8	AF352735	AF352735 Forsythia
2	260.8	31.8	933	8	NTAJ3177	AF223177 Nicotiana
3	260.4	31.8	930	8	AF286650	AF286650 Cucumis s
4	260.4	31.8	930	8	AF286651	AF286651 Cucumis s
5	252.6	30.8	1089	8	AF072447	AF072447 Ipomoea t
6	252.6	30.8	1152	8	AF072447	AF223178 Nicotiana
7	250.4	30.6	1064	8	NTAJ3178	D88121 Vigna ungu
8	246.2	30.1	1179	8	D88121	AF072449 Ipomoea t
9	244.4	29.8	3021	8	AF070221	AF070221 Ipomoea t
10	243	29.7	1120	8	AF072448	AF072448 Ipomoea t
11	243	29.7	1144	8	AF072450	AF072450 Ipomoea t
12	223.2	27.3	992	8	AB018559	AB018559 Citrullus
13	206.4	25.2	834	8	AF352734	AF352734 Podophyll
14	201.8	24.6	1176	8	AF099603	AF099603 Arabidops
15	200.8	24.5	75948	8	AC037424	AC037424 Arabidops
16	200	24.4	107109	2	OSJN00137	AL662936 Oryza sat
17	200	24.4	107109	2	OSJN00145	AL662933 Oryza sat
18	188.8	23.1	980	8	TOBTFHP1	D29976 Tobacco mRN
19	188.8	23.1	1025	6	E09193	E09193 Transcripti
20	187	22.8	111844	2	OSJN00288	AL731639 Oryza sat
21	179.8	22.0	922	8	LES277945	AL731639 Oryza sat
22	175.8	21.5	2226	8	LES277945	AB070220 Ipomoea t
23	170.8	20.9	79867	8	ATT18N14	AL132968 Arabidops
24	170.6	20.8	884	8	LEU21801	U21801 Lycopersico
25	164	20.0	90864	8	AP004949	AP004949 Lotus jap
26	162.8	19.9	944	8	AF097651	AF097651 Pisum sat
27	159.6	18.5	969	8	AF053638	AF053638 Pisum sat
28	156.4	19.1	780	8	AJ345026	AJ345026 Digitalis
29	155	18.9	1050	8	AF370319	AF370319 Arabidops
30	153.4	18.7	106329	8	AC004411	AY063106 Arabidops
31	153.4	18.7	106329	8	AC004411	AY063106 Arabidops
32	150.2	18.3	774	6	AX412402	AX412402 Sequence
33	150.2	18.3	774	6	AX412739	AX412739 Sequence
34	150	18.3	77401	8	AB026657	AB026657 Arabidops
35	147.8	18.0	143407	8	AF003825	AF003825 Oryza sat
36	135.8	16.6	817	8	AF053639	AF053639 Pisum sat
37	130.8	16.0	999	8	AY052216	AY052216 Arabidops
38	127.8	15.6	911	8	SLU53828	U53828 Silene lati
39	126	15.4	151914	2	AP004333	AP004333 Oryza sat
40	119.6	14.6	929	8	SLU53829	U53829 Silene lati
41	118.8	14.5	354	8	AF188188	AF188188 Cucumis s
42	118	14.4	1124	8	SLU53827	U53827 Silene lati
43	114.8	14.0	1187	6	AR007559	AR007559 Sequence
44	113.6	13.9	153308	8	AC118668	AC118668 Genomic s
45	109.4	13.4	163154	2	AC092557	AC092557 Oryza sat

ALIGNMENTS

RESULT 1
AF352735
LOCUS
DEFINITION
Forsythia x intermedia stem secoisolariciresinol dehydrogenase
mRNA, partial cds.
831 bp mRNA linear PLN 23-APR-2001
ACCESSION
AF352735
VERSION
AF352735.1 GI:13752457
KEYWORDS
Forsythia x intermedia.
Forsythia x intermedia
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Lamiales; Oleaceae; Forsythia.
REFERENCE
1 (bases 1 to 831)
AUTHORS
Xia,Z.Q., Costa,M.A., Pelissier,H.C., Davin,L.B. and Lewis,N.G.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
99164-6340, USA

Secoisolariciresinol dehydrogenase purification, cloning, and functional expression. Implications for human health protection J. Biol. Chem. 276 (16), 12614-12623 (2001)

2 (bases 1 to 831)
Xia, Z.Q., Costa, M.A., Davin, L.B. and Lewis, N.G.
Direct Submission
Submitted (26-FEB-2001) Institute of Biological Chemistry, Washington State University, 461 Clark Hall, Pullman, WA

FEATURES
Source
Location/Qualifiers
1..831
/organism="Forsythia x intermedia"
/db_xref="taxon:55183"
1..>831
/note="SDH-F1321; NAD-dependent; catalyzes enantiospecific conversion of (-)-secoisolariciresinol into (-)-matairesinol"
/codon_start=1
/product="stem secoisolariciresinol dehydrogenase"
/protein_id="AAK38665.1"
/db_xref="GI:13752458"
/translation="MAATSQVLTATARLEKGLVALITGGASGIGETAKLFQSHQAKV AIADVDLGHVSVEAIGTSNSYTHCDVNDGKVNADNTYSTYKGLDLMFNSAGI SDNRPRIIDNEKADPERVFSNVTVGFCLMKHAARVIMPARSGNIISTASLKTMGG GSHAYCGKSHKAVLGRNLRLAYVELGFGIRVNCSPFLPTALGKFKSGIKNEEFEN VINFAGNLGPKRFNVEDVANAALYLASDEAKYVSGHNLFDGGFSVCNSVIKVFQYD S"

BASE COUNT 244 a 169 c 192 g 226 t

ORIGIN

Query Match 98.5%; Score 806.4; DB 8; Length 831;
Best Local Similarity 99.3%; Pred. No. 4.3e-195;
Matches 810; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CAGCTTGGAAGTGCATCGCAAGAGCTGAGAGAAAGTTCCCTTTATACAGAGGA 63
DB 16 CAGGTTCTACTGATCGCAAGAGCTGAGAGAAAGTTCCCTTTATACAGAGGA 75
QY 64 GCCAGTGAATTTGAGAACACACAGAAACTCTTCTCCACATGGAGCCCAAGTTGCC 123
DB 76 GCCAGTGAATTTGAGAACACACAGAAACTCTTCTCCACATGGAGCCCAAGTTGCC 135
QY 124 ATTGCTGATGTCACAGATGAATTAGTCTACTGATGTTCTGAGGCCATTTGCCAAT 183
DB 136 ATTGCTGATGTCACAGATGAATTAGTCTACTGATGTTCTGAGGCCATTTGCCAAT 195
QY 184 TCCACCTACATCCACTGTGATGTTACTTAATGAAGCGGTGTTAAAAATCCCGTGGACAAC 243
DB 196 TCCACCTACATCCACTGTGATGTTACTTAATGAAGCGGTGTTAAAAATCCCGTGGACAAC 255
QY 244 ACAGTTTCAACCTATGGAACACTGGACATTTATGTTGAGCAATGCAGGAATTTCTGATCCC 303
DB 256 ACAGTTTCAACCTATGGAACACTGGACATTTATGTTGAGCAATGCAGGAATTTCTGATCCC 315
QY 304 AACAGGCCCGCATCATAGCAACGAAAGACAGATTTGACGGGTCTCGAGTGAAT 363
DB 316 AACAGGCCCGCATCATAGCAACGAAAGACAGATTTGACGGGTCTCGAGTGAAT 375
QY 364 GTAACCGAGTTTTCCTATGCATGAGCAGCAGCAGCAGTGTATGATCCACAGCAGT 423
DB 376 GTAACCGAGTTTTCCTATGCATGAGCAGCAGCAGCAGTGTATGATCCACAGCAGT 435
QY 424 GGCAACATAATTTCCACTGCTAGTTTAAGCTCAACTATGGGTGGTGTCTTCATATGCC 483
DB 436 GGCAACATAATTTCCACTGCTAGTTTAAGCTCAACTATGGGTGGTGTCTTCATATGCC 495
QY 484 TATTGTGGTTCACAGCATGTTGTGTTAGGCTTTACTAGGAATCTGGCAGTCGAGCTCGGA 543
DB 496 TATTGTGGTTCACAGCATGTTGTGTTAGGCTTTACTAGGAATCTGGCAGTCGAGCTCGGA 555
QY 544 CAATTTGGSCATTAGGGTTAATGTTGTTGTCCTTCCTTCGGGCTTCCCTAGCGCTTTAGGCAAG 603

Db 556 CAATTTGGCATTTAGGGTTAATGTTTGTCTCCTTCGGGCTTCCTACGGCTTTAGGCAAG 615
QY 604 AAATTTTCAGGGATTAAAAATGAAGAAGAAATTTGAGAATGTAAATAAACTTTTCGCGGAAAT 663
Db 616 AAATTTTCAGGGATTAAAAATGAAGAAGAAATTTGAGAATGTAAATAAACTTTTCGCGGAAAT 675
QY 664 TTGAAAGGTCCAAAATTTAATGTTGAGGATGTTGCAATGCGAGCTCTTTATCTGGTAGT 723
Db 676 CTGAAAGGTCCAAAATTTAATGTTGAGGATGTTGCAATGCGAGCTCTTTATCTGGTAGT 735
QY 724 GATGAGGCAAAATACGTGAGTGACACAAATCTTCAATGATGAGGCTTCAGGCTCGC 783
Db 736 GATGAGGCAAAATACGTGAGTGAGTGACACAAATCTTCAATGATGAGGCTTCAGGCTCGC 795
QY 784 AATTCGTATCAAAAGTTTCCAAATATCCAGATTCCT 819
Db 796 AATTCGTATCAAAAGTTTCCAAATATCCAGATTCCT 831

RESULT 2
LOCUS NTAUJ3177 933 bp mRNA linear PLN 08-FEB-1998
DEFINITION Nicotiana tabacum mRNA for short chain alcohol dehydrogenase.
ACCESSION AJ223177
VERSION AJ223177.1 GI:2739278
KEYWORDS short chain alcohol dehydrogenase.
SOURCE Nicotiana tabacum.
ORGANISM Nicotiana tabacum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 933)
Moenke, G.
Direct Submission
Submitted (17-DEC-1997) Moenke G.; Molecular Genetics, Institut fuer Pflanzen-genetik und Kulturpflanzenforschung, Gatersleben, D-06466, GERMANY
2 (bases 1 to 933)
Herbers, K., Monke, G., Badur, R. and Sonnwald, U.
A simplified procedure for the subtractive cDNA cloning of photoassimilate-responder genes: isolation of cDNAs encoding a new class of pathogenesis-related proteins
Plant Mol. Biol. 29 (5), 1027-1038 (1995)
96145513
855446
Location/Qualifiers
1..933
/organism="Nicotiana tabacum"
/variety="SNN"
/db_xref="taxon:4097"
/tissue_type="leaf"
5..859
/codon_start=1
/product="short chain alcohol dehydrogenase"
/protein_id="CAA11153.1"
/db_xref="GI:2739279"
/translation="MASYSLSPIAKRLKLEKVALITGGASGIGAAATRLFVQHKAVT IADIDNLTSLVQEIQNEHTIFVHCNVAVESDVQNVVDATIAKFGKLDIMFNSAGIG KSTISLIDVDYDIKTVDVNYFAFCFAKHAARVIMPFKKSIIPTASAATVSGI VPHAYSASKGAVLGFSGKNGVELGKGVKNCVSPHYISTPLVNLALGAERIAEKW FAAGNLKGLALLDEEVAKAVLYLASDSDSKYVSGMNLVIDDGFSTTNVALTEAYNKL F SSENTHQC"
497..529
misc_feature
/note="family signature"
BASE COUNT 294 a 162 c 214 g 263 t
ORIGIN

Query Match 31.8%; Score 260.8; DB 8; Length 933;
Best Local Similarity 57.9%; Pred. No. 8.2e-56;
Matches 463; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

QY	7	CTTCGAACTGCATTCGCAAGAGGCTAGAGAGAAAGTTGCCCTTATACAGGAGGACC	66
Db	20	CTTCCATACCAATAGCAAAAGGCTAGAGTAAGTAGCATAAATAGTGTGGAGCT	79
QY	67	AGTGGAAATTCGAGAAACACAGACAAACACTCTCTCCCAACATGAGCCAAAGTTGCCATT	126
Db	80	AGCGCATAGGACAGCCAGCTAGCTCTTTTGTTCACATGGTCAAGGTTTACAAAT	139
QY	127	GCTGATGTCGAAGATGAATAGTCACTCAGTTGTGCGAGCCCATTTGCCACTTCCAAATTC	186
Db	140	GCAGACATTCAAGACACACCTTGGAACTCTTTTAGTACAAGAAATTCGCAACGAACATACA	199
QY	187	ACCTACATCCACTGCTGATTTACTATATGAAGAGCGGTGTTAAATAATGCCGTGGACACACA	246
Db	200	ATCTTTGTCATTCGCAATGTGCTGTGTATGATCAGACGTTCNAAATGTTAGATGCAACA	259
QY	247	GTTCCTAACCTATGGAACACTGGACATTAATGTTCCAGCAATGCGAGAAATTTCTGATCCCAAC	306
Db	260	ATTGCCAAATTTGGTAGCTGGACATAATGTTTCAGTAACGCTGGTATAGGAGGTAAGTCA	319
QY	307	AGGCCCGGCATCATAGACACACGAAGAGACACTTTTGAACGCGTCTCTAGTGTGTAATGTA	366
Db	320	ATTTCAGCATCTTAGATGTCGATTCAGACATAATTAAGACAGTGTTCGATGTAAACATT	379
QY	367	ACCGGAGTTTTCATGATGATGAAGCAGCAGCAGCTGTTATGATTCAGCAGCCAGTGGC	426
Db	380	GTTCGGCGCTCTTTTTCGGCTAAGCATGTGCTAGAGTAATGATTCATTCAGAAAGGT	439
QY	427	AACATAATTTCCACTGCTAGTTTAAAGCTCAACTATGGTGGTGGTTCACATGCTCAT	486
Db	440	TCCATTAATCTTACACAGAGTCTGCGACAGTGGTCTCTGGTATTTGTCGCCGACGCATAT	499
QY	487	TGTGTTTCAAAGCATGCTGTGTAGCCCTTACTPAGGAATCTGCGAGCTCGAGCTCGGACAA	546
Db	500	TCGTCATCAAAAGTGCAGTTTGGGATTTTCCAGACATTTGGGTGCGAATTAGGGAAG	559
QY	547	TTTGGCATAGGTTTAATGTTGTCTCTCTTCGGGCTTCCTACGGCTTTAGGCAAGAAA	606
Db	560	TACGGATAAAAGTTAACTGTGTTTCTCCTCAATTACATAGCACACACTTGTATTGAAT	619
QY	607	TTTTTCAGGATTAATAATGAAGAGAAATTTGAGAATGTAATAACTTTGCGGAAATTTG	666
Db	620	GCCTTTGGAATAGCTGAGAGAGATAGCGGAAATAATGTTTCCAGCAGAGGAGAAATTTG	679
QY	667	AAAGGTCCAAATTAATGTTGAGGATGTTGCCAATGCAGCTCTTATCTGCGTGTAGTAT	726
Db	680	AAAGGAGCTTTACTAGATGAAGAGGAGTTCGCAAGGCGAGTGTCTATACATTAGCAAGTAT	739
QY	727	GAGGCAAAATACGTGATGGACACATCTGTTCAATGATGGAGGTTTCAGCGTCTGCAAT	786
Db	740	GATTCATAATATGATGATGGTATGAATCTGGTTATTGATGGTGGTGTATTAGTACCACAAAT	799
QY	787	TCGTGAATCAAGTGTTCCTCA	806
Db	800	GTGGCTTTAAGAGGCGCTA	819
RESULT	3		
AF286650			
LOCUS	AF286650	930 bp	mRNA linear PLN 02-AUG-2001
DEFINITION	Cucumis sativus CTA (CTA) mRNA, complete cds.		
ACCESSION	AF286650		
VERSION	AF286650.1	GI:15077027	
KEYWORDS			
SOURCE	Cucumis sativus.		
ORGANISM	Cucumis sativus.		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.		
REFERENCE	1 (bases 1 to 930)		
AUTHORS	Sun, J.-Q., Li, Q.-Z., Li, X.-G. and Zhang, X.-S.		
TITLE	Gene cloning and expression of CTA in cucumber		
JOURNAL	Unpublished		

REFERENCE	2 (bases 1 to 930)		
AUTHORS	Li, Q.-Z., Li, X.-G., Sun, J.-Q. and Zhang, X.-S.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-JUL-2000) College of Life Sciences, Shandong Agricultural University, 61 Daizong Street, Taitan, Shandong 271018, P.R. China		
FEATURES			
source	1..930		
gene	/organism="Cucumis sativus"		
	/db_xref="taxon:3659"		
CDS	1..930		
	/gene="CTA"		
	22..837		
	/note="short chain alcohol dehydrogenase; similar to TASSELSEED2"		
	/codon_start=1		
	/product="CTA"		
	/protein_id="AAK83035.1"		
	/db_xref="GI:15077028"		
	/translation="MSIQLLPAAIRRLKGVAVITGGARGIGEQTAKLFFKHAKVVI		
	ADIODHLGQTLCKDLGSSVDFVHCDVTKEKDVETAVDTVAISKYGLDMLNAGVFE		
	ESPFDILKDDPLTFQVVNVLVGASLGRHAARVMPKAGRSIVTPTASICVIGGI		
	GTHAYTSSKHVGLGMRNAVDLGRYIRVNCVSPNVPTENGRLKFLKVDGGEPPSF		
	YWSLKNGLDILREEDVEAVVILGSDSCVSLNLIYDGGFTVFNQALCSFRS"		
BASE COUNT	269 a 160 c 251 g 250 t		
ORIGIN			
Query Match:	31.8%; Score 260.4; DB 8; Length 930;		
Best Local Similarity	60.0%; Pred. No. 1e-55;		
Matches	473; Conservative 0; Mismatches 306; Indels 9; Gaps 2;		
QY	1	ATGCAGCTTCGAACTCGCAAGAGGCTAGAGAGAAAGTTGCCCTTATAACAGGA	60
Db	28	ATCCAATTCCTCCGCCCATTCGCAAGAGACTTGAAGGTAAGTAGCCGTAACTACGTGT	87
QY	61	GGAGCAGTGGAAATGGAGAAACCCAGCAAAACTCTTCTCCCAACATGGAGCAAAAGTT	120
Db	88	GGGCTAGAGGAATTTGGGAAACAAACAGCAAGCTCTCTTCAAGCATGGAGCAAAAGT	147
QY	121	GCCATTGCTGATGCCAAGATGAATTAGGTCACTCAGTTGTCGAGGCCATTGGCACTTC	180
Db	148	GTATTTCAGACATTCAGACACCATTTAGGTCAAACTACTATGAAGGATCTCGGTCACTCA	207
QY	181	AATTCCACCTACATCCACTGCTGATGTTACTTAATGAAGAGCGGTGTTAAATAATGCCGTG	240
Db	208	TCTTCT	267
QY	241	AACACAGTTTCAACCTATGGAAACTGGACATTTATGTTACGCAATGCAGGAATTT---CT	297
Db	268	ACGGCAGTCTCCAAAGTACGGAAGTGTAGACATCATGTTGAACAATGCAGGAGTTTTCGAG	327
QY	298	GATCCCAACAGGCCCGCATATAGACAAGAAAGAGACATTTGAACCGCTTCTCAGT	357
Db	328	GAATCTCCAAACTTCGACATTCCTGAAGATGATCCATTAACCTTTTCAGAGATGGTGAAC	387
QY	358	GTAATGTAACCGGAGTTTTCCTTATGCATGAAGCAGCAGCAGCTGTATGATTTCCAGCA	417
Db	388	GTCAACCTGGTTGGGGCTCTCTCGGAACGAGACACGACGACGAGTAAATGAACACGCG	447
QY	418	CGCAGTGGCAACAATAATTTCCACTGCTAGTTTAAAGTCAACTATGGGTGGTGTCTTCA	477
Db	448	GGTCGAGGAGCATCTGACGACGAGGATATATGTCGGTGTATGGGGGAATTTGGGACG	507
QY	478	CATGCCATTTGTTGTTCAAGCATGCTGTTAGCCCTTACTAGGAATCTGGCAGTCGAG	537
Db	508	CATGCTTATACAAGCTCGAAGCATGGGTGTGGGATGATGAGGAATGCAGCTGGAT	567
QY	538	CTCGGACAAATTTGGCATTAGGGTTAATTTGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	597
Db	568	TTGGGAGATATGGGATAGGGTAAATTTGTTTTCACCAATATGATGCCAACTGAATG	627
QY	598	GGCAAGAAATTTTCAGGGATTAATAAATGAAGAAGATTTTGAGAATGTAATAAATTTTGGC	657

```

Db 628 GGAAGAAAGTGTTCAGAGTTAAAGATGGTGGGGAATTTCCAAAGTTT-----CTACTCG 681
QY 658 GGAATTTTGAAGGTCCTCAAAATTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTG 717
Db 682 AGTTTAAAAAATGAGATATTTAAAGGAAGAAGATGTGGTGAAGCTCTTGTGTATTG 741
QY 718 GCTAGTGTATGAGGAAAATACGTGAGTGGACAACTCTGTTCAATTGATGGAGGTTTACG 777
Db 742 GGGAGTGCAGAGTCAAAGTGTGTGAGTGGACTCAACTTGATTGTTGATGGAGGCTTTACT 801
QY 778 GTCGTCAA 785
Db 802 GTTGTCAA 809

RESULT 4
AF286651
LOCUS Cucumis sativus TASSELSEED2-like protein (CTS2) mRNA, complete cds.
ACCESSION AF286651
VERSION AF286651.1 GI:15077029
KEYWORDS Cucumis sativus.
SOURCE Cucumis sativus.
ORGANISM Cucumis sativus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
1 (bases 1 to 930)
Sun, J.-Q., Li, Q.-Z. and Zhang, X.-S.
Functional analysis of CTS2 in cucumber
2 (bases 1 to 930)
Sun, J.-Q., Li, Q.-Z. and Zhang, X.-S.
Direct Submission
TITLE Submitted (13-JUL-2000) College of Life Sciences, Shandong
JOURNAL Agricultural University, 61 Daizong Street, Taian, Shandong 271018,
P.R. China
FEATURES
source
1..930
/organism="Cucumis sativus"
/db_xref="taxon:3659"
1..930
/gene="CTS2"
22..837
/gene="CTS2"
/codon_start=1
/product="TASSELSEED2-like protein"
/protein_id="AK83036.1"
/db_xref="GI:15077030"
/translation="MSIQLLPAIARRLEGVAVITGGARGIGETAKLFFKHGAKVVI
ADIQHLGOTLCKDLQSSVFEVHCDVTKEKDVETAVDTAVSKYKLDIMLNAGVFE
ESNFDLKDPLITQVVVNLVGLFGRKHAARVMKPAGRGSIIVTASICSIVIGI
GTHAYTSSKHVGLMRNAVDLGRYVNCVSPNVVPTVMGRKLFKVDGGEFPFSF
YWSLKNGLDILREEDVGEAVYLGSDSKVCSGLNLIVDGGFTVVNQALCSFRS"
BASE COUNT 270 a 158 c 249 g 253 t
ORIGIN
Query Match 31.8%; Score 260.4; DB 8; Length 930;
Best Local Similarity 60.0%; Pred. No. le-55;
Matches 473; Conservative 0; Mismatches 306; Indels 9; Gaps 2;
QY 1 ATGCAGCTCGAAGTTCGATTCGCAAGAAGCTTGAAGAGAAAAGTTGCCCTTATACAGGA 60
Db 28 ATCCANTGTCTCCCGCATTCGCAAGAAGCTTGAAGGTAAAGTAGCGTAAATCACTGGT 87
QY 61 GGAGCGATGGAATTGGAGAAACACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTT 120
Db 88 GGGCTAGAGGAATTTGGGGAACAAAGCGAAGCTCTTCTTCAAGCATGGAGCCAAAGTG 147
QY 121 GCCATTGCTGATGCCAAGATGAATTAGTCTACTAGTTGTCTCGAGGCCATTGGCACTTCC 180
Db 148 GTTATTGCAGACATTCAAGACCACTTTAGGTCAAACTATGTAAGGATCTCGTCAGTCA 207

```

```

QY 181 AATTCACCTACATCCACTGATGTTTACTAATGAAGACGGTGTAAAAATGCCGTGAC 240
Db 208 TCTTCTGTCTTCGTTCAITTCGCACTAAACAAAGAGAACGTTGAACTCGCGTTGAC 267
QY 241 AACACAGTTTCAACCTATGGAAGAACTGGACATTTATGTTACAGCAATGCAGAAATTT--CT 297
Db 268 ACGGCAGTCTCCAAGTACGGAAGTTAGACATCATGTTGAACATTCAGAGTTTTCGGA 327
QY 298 GATCCACAGAGCCCGCATCATAGACAAACGAAAACAGACACTTTGAACCGTTCACGT 357
Db 328 GAATCTCCAAACTTCGACTTTCTGAAAGATGTCCATTACCTTTACAGAGTGGTGAAC 387
QY 358 GTAATGTAAACCGAGTTTCTTATGCAATGAAGCAGCAGACAGTGTATGATTCACGA 417
Db 388 GTCAACCTGGTTGGGGCCCTTTCTCGGAACGAAACACGACGACGAGTAATGAACCCAGC 447
QY 418 CGCAGTGGCAACATAAATTTCCACTGCTAGTTTAAAGCTCAACTATGGGTGGTCTTTCA 477
Db 448 GGTGAGGAGGAGCATGTCGACGACAGCAGATATATCTCGGTGATTGGGGAAATTTGGGACG 507
QY 478 CATGCCATATTGTTGTTCAAAACATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCGAG 537
Db 508 CATGCTTATACAAGCTCGAAGCATGGGGTGTGGGATTGATGAGGAATGCAGCTGGAT 567
QY 538 CTCGGACAAATTTGGCATTAGGTTAATGTTGTTGCTCTCTTCGGGCTCTACGCGCTTTA 597
Db 568 TTGGAAGATATGGGTTAGGTTAAATTTGTTTTCACCAATGTAGTGCCCACTGAATG 627
QY 598 GGCAGAAATTTTCAGGGATAAATAATGAAGAAGATTGAGAATGTAATAAATCTTTGCG 657
Db 628 GGAAGGAAGTTGTTCAAGGTTAAAGATGGTGGGAAATTTCCAAAGTTT-----CTACTGG 681
QY 658 GGAATTTGAAGAGTCCAAATTTAATGTTGAGGATGTGCCAATGCAGCTCTTTATCTG 717
Db 682 AGTTTAAAAATGGAGATATTTTAAGGGAAGAAGATGTGGGTGAAGCTGTGTGTATTG 741
QY 718 GCTAGTGTATGAGCAAAATACGTGAGTGGACACAACTGTTTCATTGATGAGGAGTTTCA 777
Db 742 GGGAGTGCAGAGTCCCAAGTGTGTGAGTGACTCAACTTGATTGTTGATGAGGCTTTACT 801
QY 778 GTCGTCAA 785
Db 802 GTTGTCAA 809

RESULT 5
AF072447
LOCUS Ipomoea trifida short-chain alcohol dehydrogenase (SSP) mRNA, SSP-1
DEFINITION Ipomoea trifida short-chain alcohol dehydrogenase (SSP) mRNA, SSP-1
ACCESSION AF072447
VERSION AF072447.1 GI:3598856
KEYWORDS ipomoea trifida.
SOURCE ipomoea trifida.
ORGANISM ipomoea trifida.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
1 (bases 1 to 1089)
Koyama, Y., Kadota, N., Sassa, H., Kakeda, K., Tsuchiya, T.,
Ishimizu, T., Kondo, K. and Norioka, S.
Self-incompatibility plant, Ipomoea trifida, is a member of
short-chain alcohol dehydrogenase family
Unpublished
2 (bases 1 to 1089)
Koyama, Y. and Kadota, N.
Direct Submission
TITLE Submitted (18-JUN-1998) Faculty of Bioresources, Mie University,
Kamihama-cho, Tsu, Mie 514-8507, Japan
FEATURES
Location/Qualifiers
1..1089
source

```



```
/organism="Ipomoea trifida"
/db_xref="taxon:35884"
/note="wild diploid species related to sweet potato"
1. .1089
/gene="SSP"
/allele="SSP-1"
44. .898
/gene="SSP"
/note="self-incompatibility-locus specific stigma protein"
/codon_start=1
/product="short-chain alcohol dehydrogenase"
/protein_id="AAC35340.1"
/db_xref="GI:3598857"
/translation="MAASLSQLPGSKRLGKVAITGAANGIGATTARLFARHQGKVI
IADIDKNGHSAEIGPEYALFICHDVIESDVQNAVDTTVSRVKGKLDIMFSNAGVA
GSRDTSILEASPENLVFETNVGAFFCAKHAARVMPARKSGVIFASASAEEVFGI
TSDTVAASKAVVGLKSLICVMKGYKIKANCVSPVILTKLGMSIMPTQDPKLAEEI
VAEASNFGKLTITDEVAEALYLAGDESKFVSGNLNLLIDGGFTTNIAFQVAVEKVL
GGGEETTA"
BASE COUNT 303 a 249 c 282 g 255 t
ORIGIN
Query Match 30.8%; Score 252.6; DB 8; Length 1089;
Best Local Similarity 57.4%; Pred. No. 1e-53;
Matches 456; Conservative 0; Mismatches 339; Indels 0; Gaps 0;
QY 7 CTTCTGAACCTGCTTCGCAAGAGGCTAGAGAAAGTTGCCCTTATAACAGGAGGAGCC 66
Db 59 CTTCAACTCCCGGGTCCAAAGACTAGAGGAAGGTTGCTATATAACAGCGCGAGCC 118
QY 67 AGTGGAATGGGAACACACAGCAAACTCTCTCCCAACTGAGGCCAAAGTTGCCATT 126
Db 119 AACGGCATAGGACCAACCGCGAGGCTCTTCTCAACATGGCTGCAAGGTTATCAT 178
QY 127 GCTGATGTCGAAGTAAATAGTCTACTCAGTTGTGAGGCCATTGGCACTTCCATTCC 186
Db 179 GCTGATGTCGAAGTAAATAGTCTACTCAGTTGTGAGGCCATTGGCACTTCCATTCC 238
QY 187 ACCTACATCCACTGTGATGTTACTAATGAAGAGCGGTGTTAAATAATGCCGTGGACACACA 246
Db 239 CTATTCATCCACTGCCAGCTCAGAACTGCAATCAGAGCTCCAAATCGGTGGACACACC 298
QY 247 GTTTCACCTATGGAAATCGGACATATTGTTCCAGCAATCGAGGAATTTCTGTATCCCAAC 306
Db 299 GTTTCAGGTTACGGAAGTTAGACATCATGTTCCAGCAATCGAGGCGTCGCGAGGAGCGC 358
QY 307 AGCCCGCGCATCATACACAGCAAAAGCAGACTTTGAACGCGTTCTCAGTGTAAATGTA 366
Db 359 GACACAGCATCTTGGAGCGCCAGCCAGAAAACATCAACCTAGTTTTCGAGAGCAATGTG 418
QY 367 ACCGAGGTTTTCTATGTCATGAAGCAGCAGCAGCGTGTATGATTTCCAGCAGCAGCTGGC 426
Db 419 TTCGAGCATCTTCTGCGCCAAACACGCGGCAAGGTTGATGATCCGCGAGGAGGCG 478
QY 427 ACATATATTCCTACTGCTAGTTTAAGCTCAACTATGGGTGGTGGTTCCTTCAATGCTTAT 486
Db 479 AGCGTAATATCTCCGAGCGCGCGCTCGGAGGTTCTCGGATCACTCCGACAGCTAC 538
QY 487 TGTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGCAGCTCGAGCTCGGACAA 546
Db 539 ACGGCATCAAGTGCAGCGCTGGGAGCTATGCAAGAGGCTGTGCGTTGAAATGGGAAA 598
QY 547 TTTGGCATTAGGTTAATGTTGTCTCTTCGCGGCTTCTTACGCTTTTAGCAAGAAA 606
Db 599 TATGGATCAAGCCAACTTCGCTTTCGCTTATGTAATTTCTGACTAAGTTGGGAATGAGT 658
QY 607 TTTTCAGGATTAAATAAGCAAGAAATTTGAGAATGTATAAATTTTTCGGGAAATTTG 666
Db 659 ATAATGCCCAACCCAGGATAGGAAATTTGGCGGAGGAAATTTGTGCGAAGCTTCGAATTT 718
QY 667 AAGGTCCAAATTTAATGTTGAGGATGTTGCCAATGCAAGCTTTTATCTGGCTAGTGAT 726
Db 719 AAAGGGAAACTTTGACGACAGAGGATGATGACAGAGCGGCTTTGTACCTCGCGCGGAC 778
```

```
QY 727 GAGGCAAAATACGTGAGTGACACAAATCTTTCATTGATGAGGTTTCAGCGTCTCAAT 786
Db 779 GAGTCTAAGTTCGTAAGCGGCTCAACCTTTGATTGATGGGGATTTACCACCAAT 838
QY 787 TCTGTAAATCAAGTG 801
Db 839 ATAGCCTTTCAAGTG 853
RESULT 6
NTAJ3178 Nicotiana tabacum SCANT gene. 1152 bp DNA linear PLN 30-JAN-1998
DEFINITION AJ223178
ACCESSION AJ223178
VERSION AJ223178.1 GI:2791347
KEYWORDS SCANT gene; short chain alcohol dehydrogenase.
SOURCE Nicotiana tabacum.
ORGANISM Nicotiana tabacum
REFERENCE 1 (bases 1 to 1152)
AUTHORS Herbers,K., Monke,G., Badur,R. and Sonnwald,U.
TITLE A simplified procedure for the subtractive cDNA cloning of
photoassimilate-responder genes: isolation of cDNAs encoding a new
class of pathogenesis-related proteins
JOURNAL Plant Mol. Biol. 29 (5), 1027-1038 (1995)
MEDLINE 96145513
PUBMED 8553446
REFERENCE 2 (bases 1 to 1152)
AUTHORS Moenke,G.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Moenke G., Molecular Genetics, Institut
fuer Pflanzenzuenetik und Kulturpflanzenforschung, Gatersleben,
D-06466, GERMANY
FEATURES
Source
1. .1152
/organism="Nicotiana tabacum"
/variety="SNN"
/db_xref="taxon:4097"
/tissue_type="leaf"
Join(5. .43,134. .949)
/gene="SCANT"
Join(5. .43,134. .949)
/gene="SCANT"
Join(5. .43,134. .949)
/gene="SCANT"
/codon_start=1
/product="short chain alcohol dehydrogenase"
/protein_id="CAAL1154.1"
/db_xref="GI:2791348"
/db_xref="SPTREMBL:O50038"
/translation="MASYSLPSPKARLEKGVKALITGGASGIGATARLFVQHGAKVT
IADIDNGLTSLVQIEGNETIFVHCNVAVESDVQNAVDTTVSRVKGKLDIMFSNAGIG
KSISSILDVDYDIITKTFVDNIVGAFFCAKHAARVMPARKSGVIFASAAVVSIGI
VPHASGKAVLGFSKNGIKVNCVSPHYISTPLVLNAGTAEREIAEKW
FAACGNLKGALLDEEVAKANVLYLASDDSKYVSGNMLYIDGGFTTNVATLTAENKLF
SSENTHQC"
<5. .43
/gene="SCANT"
/number=1
44. .133
/gene="SCANT"
/number=1
134. .949
/gene="SCANT"
/number=2
587. .619
/gene="SCANT"
/number=2
/feature="family signature"
BASE COUNT 356 a 196 c 244 g 356 t
ORIGIN
```



```

REFERENCE
AUTHORS Tsuchiya,T., Suwabe,K., Watabe,T., Kagaya,Y. and Kowiyama,Y.
TITLE Characterization of promoter region of the SSP gene from Ipomoea
JOURNAL tritrida regulates stigma-specific expression
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 3021)
TITLE Tsuchiya,T.
JOURNAL Direct Submission
SUBMITTED (15-AUG-2001) Tohru Tsuchiya, Mie University, Lab. of
Plant Molecular Genetics and Breeding; Kamihama, 1515, Tsu, Mie
514-8507, Japan (E-mail:tsuchiya@bio.mie-u.ac.jp,
Tel:81-59-231-9515, Fax:81-59-231-9515)
FEATURES
source Location/Qualifiers
1..3021
/organism="Ipomoea trifida"
/strain="H77-2"
/db_xref="taxon:35884"
/clone_lib="Lambda FIXII Ipomoea trifida genomic Library"
/notes="vector:pBluescript SK-"
1543..1550
/gene="SSP10"
join(1649..1686,1787..2582)
join(1649..1686,1787..2582)
/gene="SSP10"
/codon_start=1
/product="S-Locus linked stigma protein"
/protein_id="BAB86916.1"
/db_xref="GI:19911179"
/translation="MAASSLQLPVFKRLKGVKAITGAANGIGATTARLFAHQGCKVI
IADIDKNGLSVAEIGPEYALFHCDVRIESHVQHAVDTTVSRGKLDIMFSNAGVA
GSRDTSLEASPENINLVFTNVEFGAFCAKHAARVILPARKGVSIFSAASAEVFGI
TSDYATASKAVGLKSLCVEMGKYGKIKANGCVSPVILTKLGMSTMPQDRKLAEEI
VAEASNFKGKLTATEDVAEAAALYLADGESKFSYGLNLLIDGGFTTTNTAFQAVAEKIL
G"
polyA_signal 2631..2636
polyA_signal 2775..2780
BASE COUNT 1005 a 549 g 574 g 892 t 1 others
ORIGIN

Query Match
Best Local Similarity 29.8%; Score 244.4; DB 8; Length 3021;
Matches 443; Conservative 0; Mismatches 331; Indels 0; Gaps 0;

QY 28 AGCGTAGAAGAAAGTTGCCCTTATAACAGGAGGAGCCAGTGGAAATGGAGAACACCA 87
DB 1785 AGACTAGAGGAAAGGTTGCTATCATACAGCGCGAGCCAGCGCATAGAGCAACACC 1844
QY 88 GCAAAACTCTTCCCAACTAGGAGCCAAAGTTGCCATTGCTGATGTCACAGATGAATTA 147
DB 1845 CGAGGCTCTTTGCTCAACATGCTGCAAGTCAATCTTCCATCTGCTGACATGACGACAGAAT 1904
QY 148 GGTCACTCAGTTGTCGAGGCCATTGGCACCTTCCAAATCCACTACATCCACTGCTGATGT 207
DB 1905 GGCTCTCCGTCGGGAGAGATCGGCCCGGAATACGCACTATTTCATCCACTCGGACGTC 1964
QY 208 ACTAATCAAGACGGTGTAAATAATCCGTCGGGACACACAGTTCACCTATGGAACATG 267
DB 1965 AGAATCGATCAGACTCCAAATGCGGTGACACACACACCGCTTTCCAGGTACGGAAGCTA 2024
QY 268 GACATATCTGTCAGCAATGCGAGGAATTTCTGATCCCAACAGCCCGCATATAGACAAAC 327
DB 2025 GACATCATGTCAGCAATGCGAGGCGTCGACACACACCGCTTTCCAGGTACGGAAGCTA 2084
QY 328 GAAAGACGACGTTGACGCGGTTCTCAGTGTAAATGTAACCGGAGTTTTCCTATGCAATG 387
DB 2085 AGCCGAGACACATCACTAGTTTTCGAGACGAATGTGTCGGAGCATCTTCTGCGCC 2144
QY 388 AAGCAGCAGCAGCTGTATGATCCAGCAGCGAGTGGCAACATATATTCACATGCTAGT 447
DB 2145 AAACAGCGGCGAGGTTGATGTCGCGGAGGAAGGCGAGCGTAAATATCTCGCGGAGC 2204
QY 448 TTAAGCTCAACTATGGTGGTGTCTTCCATGCTATTGTGTTCAAAGCATGCTGTG 507

```

```

Db 2205 GCGGCGTCGGAGTGTTCGGGATCACCTCGGACACGTACACGGATCCAGTGCCTC 2264
QY 508 TTAGCCCTTACTAGGAATCTGCGAGTCGAGCTCGGACAAATTTGGCATTAGGTTAATGT 567
Db 2265 GTGGGACTATGCAAGAGCCTGTGCTTGAATGGGAAATATGGGATCAAGCAACTGC 2324
QY 568 TTGTCCTCTTTCCGGCTTCTCAGCGCTTAGGCAAGAAATTTTCAGGAGTAAATAAGTAA 627
Db 2325 GTTTCGGCTTATGTAATCTGACTAAGTTGGGATGAGTATATCCCGACCCAGGATAGG 2384
QY 628 GAAGAATTTGGAATGTAATAAACTTTGCGGGAATTTCAAGGTCCAAAATTTAATGTT 687
Db 2385 AAATTGGCGGAGAAATTTGTGCGGAAGCTTCGAATTTCAAGGGAATACTTTGAGACA 2444
QY 688 GAGATGTTGCCAATGACGCTCTTTATCTGCTAGTATGAGGCAAAATACGTAGTGA 747
Db 2445 GAGATGATAGCAGAGCGCTTTGTACCTCGCGGACGAGTCTAAGTTTCGTGAGCGG 2504
QY 748 CACAATCTGTTTCATGATGAGGCTTACGCTGTCATCTGTAATCTGTAATCAAGTG 801
Db 2505 CTCACCTTCTGATTGATGGGGGATTACCAACCAATACAGCCTTTCAAGTG 2558

RESULT 10
AF072448 1120 bp mRNA linear PLN 13-SEP-1998
LOCUS Ipomoea trifida short-chain alcohol dehydrogenase (SSP) mRNA, SSP-9
DEFINITION
AF072448
ACCESSION
VERSION AF072448.1 GI:3598858
KEYWORDS
SOURCE Ipomoea trifida.
ORGANISM Ipomoea trifida.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
1 (bases 1 to 1120)
Koyama,T., Kadota,N., Sassa,H., Kakeda,K., Tsuchiya,T.,
Ishimizu,T., Kondo,K. and Norioka,S.
Self-incompatible stigma protein (SSP) from sporophytic
short-chain alcohol dehydrogenase family
Unpublished
2 (bases 1 to 1120)
Koyama,T. and Kadota,N.
Direct Submission
Submitted (18-JUN-1998) Faculty of Bioresources, Mie University,
Kamihama-cho, Tsu, Mie 514-8507, Japan
Location/Qualifiers
1..1120
/organism="Ipomoea trifida"
/db_xref="taxon:35884"
/notes="wild diploid species related to sweet potato"
1..1120
/gene="SSP"
58..891
/locus_tag="SSP-9"
/gene="SSP"
/notes="self-incompatibility-locus specific stigma protein"
/codon_start=1
/product="short-chain alcohol dehydrogenase"
/protein_id="AAC35341.1"
/db_xref="GI:3598859"
/translation="MAASSLQLPVFKRLKGVKAITGAANGIGATTARLFAHQGCKVI
IADIKNGHVAEEIGPEYALFHCDVRIESHVQHAVDTTVSRGKLDIMFSNAGVA
GSRDTSLEASPENINLVFTNVEFGAFCAKHAARVILPARKGVSIFSAASAEVFGI
TSDYATASKAVGLKSLCVEMGKYGKIKANGCVSPVILTKLGMSTMPQDRKLAEEI
VAEASNFKGKLTATEDVAEAAALYLADGESKFSYGLNLLIDGGFTTTNTAFQAVAEKIL
G"
BASE COUNT 315 a 255 c 281 g 269 t
ORIGIN

```



```

Qy 487 TGTGGTTCAAAGCATGCTGTTAGCCCTTACTAGGAATCGCAGTCGAGCTCGGACAA 546
Db 554 ACGGCATCAAGTGGCCGCTGCGGACTATGCAAGAGCTGTGCGTTGAAATGGGGAAA 613
Qy 547 TTTGGCATTAGGGTTAAATGTTTGTCTCCCTTCGGGCTTCCTACGGCTTTAGGCAAGAAA 606
Db 614 TATGGGATCAAGGCAACTGCGTTTCGCCCTTATGTAATCTGACCAAGCTGGGAATGAGT 673
Qy 607 TTTTCAGGATTAATAATGAAGAAGATTTGAGAATGTAATAAATTTGCGGGAAATTTG 666
Db 674 ATAATGCCGACCCAGGATAGAAATTCGGCGAGGAAATTTGTGCGGAAGCTTCGAATTC 733
Qy 667 AAAGTCCAAAATTAATGTTGAGGATGTTGCCAATGCAAGCTCTTTATCTGGCTAGTGAT 726
Db 734 AAAGCAAAATCTTGACGACGAGGATGTAGCAGGGCGCTTTGTACTCGCCGGCGAC 793
Qy 727 GAGGCAAAATACGTCAGTGCAGACAAATCTGTTCAATGATGAGGAGTTCAGCGTCTGCAAT 786
Db 794 GAGTCTAAGTTCGTAGCGGGCTCAACCTTTTGTATGTAAGGGGATTTACCAACCAAT 853
Qy 787 TCTGTAATCAAAGTG 801
Db 854 ATAGCCTTCAAGTG 868

RESULT 12
AB018559
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Citrullus lanatus developing seeds cDNA to mRNA.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosidae; eurosids; Cucurbitales; Cucurbitaceae; Citrullus.
1 (bases 1 to 992) Kim, J., Kang, H., Jun, S., Lee, J. and An, G.
Cloning of three gibberellin-regulated cDNAs from watermelon during
early seed development: down-regulated one cDNA and up-regulated
two cDNAs
Published Only in DataBase (2000)
2 (bases 1 to 992)
Kim, J., Kang, H., Jun, S., Lee, J. and An, G.
Direct Submission
Submitted (12-OCT-1998) Junyul Kim, Pohang University of Science
and Technology (POSTECH), Department of Life Science, Hyoja, Pohang
790-784, Republic of Korea (E-mail:flower4@postech.ac.kr,
Tel:82-562-279-5995)
Location/Qualifiers
1. 992
/organism="Citrullus lanatus"
/db_xref="taxon:3654"
/dev_stage="developing seeds"
/note="synonym:Citrullus vulgaris"
1. 992
/gene="wts2L"
65. 886
/gene="wts2L"
/codon_start=1
/product="wts2L"
/protein_id="P889230.1"
/db_xref="GI: P889230.1"
/translation="MLNRSPTPLRLDGKVALITGASGIGTSVAVRIFENGAKVII
SFSGILDVKSDELDDVSYIHCDSKEDVSNLVDAAVHRHGKLDIMYSNAGVLD
HYPAASKGAVLGLVRLNLAELGQGVNCVAPFVATGAGSRDPMQAELETVMVTT
992
WANKRVLKADIAKALYLASDDANYVSLNLVDDGVSVVNPMTLKLMD"
/gene="wts2L"
/note="27 A nucleotides"
257 a 253 c 246 g 236 t

polya_site
BASE COUNT

REFERENCE
1 (bases 1 to 834)

```

ORIGIN

```

Query Match      27.3%; Score 223.2; DB 8; Length 992;
Best Local Similarity 57.6%; Pred. No.3.2e-46;
Matches 440; Conservative 0; Mismatches 318; Indels 6; Gaps 2;

Qy 25 AGAAGGCTAGAAAGAAAAGTTGCCCTTATAACAGAGGAGGAGGAGTGGAAATGGGAAC 84
Db 95 AGAAGGCTAGAAAGAAAAGTTGCCCTTATAACAGAGGAGGAGGAGTGGAAATGGGAAC 84
Qy 85 ACAGCAAAACTCTCTCCCAACATGAGCAAAAGTTGCCATTTGCTGTGTGTCCTCAAGTAA 154
Db 155 GCAGTGGAAATTTTCCATGAAATGAGGCAAGTGAATATCGCCGATATCCAGAGCAA 144
Qy 145 TTAGGTCACTCAGTGTTCGAGGCCATTCGCACTTCCAAATTCACCTACATCCACTGTGAT 204
Db 215 ATCGGCCAAAATAATCCGACGAACCTCGCGAT---GACGTAAGCTATATCCACTCGAT 271
Qy 205 GTTACTAATGAAGACAGGTGTTAAATAATCCCTGGACACACAGTTTCAACCTATGAAAA 264
Db 272 GTGTCGAAGCAAGACAGCTCAGCAATCTGTGGACCCCGCGTGCACCGGCATGGCAAG 331
Qy 265 CTGGACATTATGTTCAAGCAATGCAAGATTTCTGATCCCAACAGGCCCGCATCATAGAC 324
Db 332 CTGGACATCATGTACAGCAACGCCGGGTTCTGACCGTTTCTTAGCGGCATATTTGGAC 391
Qy 325 AACGAAAAGCAGACTTTGAACGGGTTCTCAGTGTAATGTAACCGGAGTTTTCCTATGC 384
Db 392 GTTACAAAATCTGACTTGGACAAGGTCTTGGCGTGAACGTGATGGGAGCATTTTGGGA 451
Qy 385 ATGAAGACGCGACGACGTGTTATGATTCAGACGCGAGTGGCAACATAATTTCCACTGCT 444
Db 452 GCAAGCATGCGAGCAGAGTAATGATACCGAGAAAATGGTGCAATTTCTATTCAACAGC 511
Qy 445 AGTTTAAGCTCAACTATGSGTGGTGTCTTCTTACATGCGCTATTTGTTTCAAGCATGCT 504
Db 512 AGCGCAACCAACCAATGCTGCGCTCTCTACACCCATACGAGCGCTTCAAAATGCGCT 571
Qy 505 GTGTTAGCCCTTACTAGGATCTGAGTCGAGTCGAGTTCGAGCAATTTGGCATTAGGGTTAAT 564
Db 572 GTCTAGGGCTGTTAGGAACCTAGTCCGAGCTCGGCCAACACCGGATCAGAGTCAAC 631
Qy 565 TGTTTGTCTCTTTTCGGGCTTCTAGCGCTTTAGCAAGAAATTTTCAGGGATTAATAAT 624
Db 632 TGTGTCGCCCTTCTGTCGTCGACCGGATTCAGGG---TCCAGAGACCCGATGCAA 688
Qy 625 GAAGAAGAATTTGAGAATCTATAAACHTTTCGGGAAATTTGAAAGTCCCAAATTTAAT 684
Db 689 GCGAGGCGCTCGAGACCATGTCACACCTGGGCCAATCTCAAAGGTGCTGTCCTTAAG 748
Qy 685 GTTGGATGTTGCCAATGCAAGCTCTTTATCTGGCTAGTGATGAGGCAAAATACGTGAT 744
Db 749 GCGACGACATAGCCAGGCTCGCTCTACTTGGCTAGCGACGATGCCAACTATGTCAGC 808
Qy 745 GGACACAATCTGTTCAATGATGAGGCTTCAGCGTCTGCAATTC 788
Db 809 GGCCTTAATCTTGTGGTCGATGAGGCTACAGTGTGTGTCATCC 852

RESULT 13
AF352734
LOCUS
DEFINITION
Podophyllum peltatum rhizome secoisolariciresinol dehydrogenase
mRNA, partial cds.
ACCESSION
AF352734
VERSION
AF352734.1 GI:13752455
KEYWORDS
Podophyllum peltatum.
Podophyllum peltatum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
Berberidaceae; Podophyllum.
1 (bases 1 to 834)

```

Qy	544	CAATTTGGCATTAGGTTAATGTTTGTCTCCTTTGGGGCTTCCTACGGCTTTAGGCAAG	603
Db	559	GAGTATGGGATCCGAGTAAACTGTGTATCTCTTATATTGTCATGCCCATTTGTTGACC	618
Qy	604	AAATTTTCAGGGATTAAAAATGAAGAACAATTCAGAAATGTAATAAACTTTTCGGGAAAT	663
Db	619	GATGTGTTGGAGTGGATTTCTAGTAGG- --TTGAGGAATTGGCACATCAAGCTGCAAAAC	675
Qy	664	TTGAAAGGTCCAAAATTTAATGTTTCAGGATGTTGCCAATGACAGCTCTTTATCTGGCTAGT	723
Db	676	CTCAAGGGACTTTGCTCAGGGCTGAGGATGTGCCGATGACGTGCGCTATTTCGCGAGG	735
Qy	724	GATGAGGCAAAATACGTGAGTGGACACAATCTGTTATGTATGAGGAGGTTTCAGCGTCTGC	783
Db	736	GATGAGTCCCAAGTATGTGAGCGGTCTGAATCTGTTATCGATGGGGGCTACACCAAACT	795
Qy	784	AATTCCTGAATC	795
Db	796	AATCCGGCTTTC	807

RESULT	14
AY099603	
LOCUS	AY099603
DEFINITION	Arabidopsis thaliana short chain alcohol dehydrogenase, putative (At1g52340) mRNA, complete cds.
ACCESSION	AY099603
VERSION	AY099603.1
KEYWORDS	FLU CDNA.
SOURCE	Arabidopsis thaliana.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsida. 1 (bases 1 to 1176)
AUTHORS	Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
TITLE	Direct Submission
JOURNAL	Submitted (24-APR-2002) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT	e-mail for correspondence: arab@sequence.stanford.edu
	RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.
	The Salk, Stanford, PCEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.
	Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.
FEATURES	Location/Qualifiers
source	1..1176 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" /clone="RAFL09-66-J21 (R24721)" /note="This clone is in pBluescript ecotype: Columbia" 1..1176 /gene="At1g52340"
gene	

CDS	122.. 979 /gene="At1g52340" /codon_start=1 /product="short chain alcohol dehydrogenase, putative" /protein_id="AAW20454.1" /db_xref="GI:20466274" /translation="MSTNTSSSYSLRFSQRLKGLVITGATGIGESIVRLPHKHG AKVYDLQDLGGEVCKSLRSGEKETAFFIHGDRVEDDISNAVDFAVKNGTILDI LNNGLGCPADIDINYSLEFEMFDVNVKGAFLSKHAARWMPKKGSLVLSLS VGVNGVGFHSVSKHVLGSTRVAELQGHGIRVNCVSPVATKIALAHLPPE EREDAFVGFNRFAANANLKGVELTVDDVANAVLFLASDDSRVISEDNLMDGGFTC TNSFVYFR"
BASE COUNT	314 a 191 c 307 g 364 t
ORIGIN	
Query Match	24.5%; Score 201.8; DB 8; Length 1176;
Best Local Similarity	56.2%; Pred. NO. 9.3e-41;
Matches 455; Conservative	0; Mismatches 327; Indels 27; Gaps 3;
QY	27 AAGCTAGAGAAAGTTGCCCTATACAGGAGGAGCCAGTGGAAATTCGAGAAACCC 86
Db	169 AAGCTTTTGGTAAAGTGGCATTCATCAGTGGAGGAGCCACAGGATAGTGGAGCAT 228
QY	87 AGCAAAACTCTCTCCACATGAGGACAAAGTTGCCATTGCTGATCTCCAGATCAAT 146
Db	229 TGTTCGCTCTCCACAGCAGCGTGCANAGTCTGATTTGATCTGCAAGATGATCT 288
QY	147 AGGTCACTCAGTTGTCGAGGCCATGGCAGCTCCCAATTCACACCTA-----CAT 194
Db	289 CGGAGGTGAGGTGTGAAAGTCTGCTTCGGTGGTGAATCCAGGAGCGCTTTTTCAT 348
QY	195 CCAGTGTGATGTACTATGATGAAGAGGTGTAAAAATGCGGTGGACACACAGTTTCAAC 254
Db	349 CCATGGCGATGTAGAGTGGAGTACATAGCAATGCGGTGATTTTGGCAGCAAAA 408
QY	255 CTATGAAAACCTGGACATATGTCAGCAATGAGCAATTTCTGATCCCAAGCCCGC 314
Db	409 TTTTGGGACGCTTGATATCTATCAATCAGGATTTATGTGAGCAGCCGTCCTCGTA 468
QY	315 CATCATACACAAACAAACGACATTTGAACCGGTCTCAGTGTAAATGTAACCGGAGT 374
Db	469 TATTCGTAATATATGTTAGTGGTTCAGATGACCTTTGATGTGATGTGAAAGGAGC 528
QY	375 TTTCCTATGATGAAGCAGCAGCGTGTATGATTCAGCAGCCAGTGGCAACATAAT 434
Db	529 TTTTCTAAGCATGAACATGCAGCTCGTGTATGATACCGGAGAGAAAGGGTCGATAGT 588
QY	435 TTCCACTGCTAGTTTAAGCTCAACTATGGTGGTGGTCTTCATGCTATTTGTGTTTC 494
Db	589 TTCCTTATGATGATGGAGGTGTTGGGAGCGGTGGTCCACATCTTATGTTGGTTC 648
QY	495 AAAGCATGCTGTAGCCCTTACTAGCAATCTGGCAGTCCGAGTCCGCAATTTGGCAT 554
Db	649 CAAGCATGCTGTAGGCTTCACTAGAGGTGTTCAGCGGAGCTTGGACGACGCGGAT 708
QY	555 TAGGGTAAATTTGTTCTCTCTTTCGGGCTTCCTAC-----GGCTTTAGCAAGAAAT 608
Db	709 AGGTGGAATCTGTTTCGCTTACGCGGTTCAGCACTAACTGCTTTGGCTCAATTTGCC 768
QY	609 TTCAGGATTTAAATGAAGAGATTTGAGATGTGATTAATAACTTTGC-----GGG 659
Db	769 GGAGGAAGAAAGACGGAGGATGCATTTGTGGTTTCAGGAATTTTGTGTCGAAACGC 828
QY	660 AAATTTGAAGGTCCTCAAAATTTATGTGAGATGTGTCATGAGCTCTTTATCTGCG 719
Db	829 GAATCTAAAGGGGTGGAACATGACGCTTATGATGATGAGCAACGCTGCTCTGTTTTCG 888
QY	720 TAGTGATGAGCAAAATACGTGATGAGCACAATCTGTTCATGATGAGGTTTCAGCGT 779
Db	889 TAGCGATGACTCGCGGTACATAGCGGAGATTAATTTGATGATGATGAGGATTCATTG 808
QY	780 CTGCAATCTGTAATCAAGTGTTCAT 808

Db	949	CACATAACCACTCCTTAAAGTCTTCAGAT	977
RESULT 15	AC037424		
LOCUS			
DEFINITION	Arabidopsis thaliana chromosome 1 BAC F19K6 bp DNA linear	75948 bp	PLN 19-JAN-2001
ACCESSION	AC037424		complete sequence.
VERSION	AC037424.10	GI:12323115	
KEYWORDS	HTG		
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE			
AUTHORS	Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J., Barnstead, M.E., Renning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Arabidopsis thaliana chromosome 1 BAC F19K6 genomic sequence.		
TITLE	Unpublished thaliana chromosome 1 BAC F19K6 genomic sequence		
JOURNAL	2 (bases 1 to 75948)		
AUTHORS	Lin, X. and Kaul, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-APR-2000)		
AUTHORS	Medical Center for Genomic Research, 9712		
TITLE	Town, C.D. and Kaul, S.		
JOURNAL	3 (bases 1 to 75948)		
AUTHORS	Submitted (19-JAN-2001)		
TITLE	Medical Center for Genomic Research, 9712		
JOURNAL	On Jan 19, 2001 this sequence version replaced gi:12280784.		
COMMENT	Address all correspondence to: at@tigr.org		
	The clone F19K6 is from Arabidopsis thaliana chromosome 1		
	The orientation of the sequence is from SP6 to 77 end of the BAC		
	Genes were identified by a combination of several methods: Gene		
	prediction programs including Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky, http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant of GlimmerW, see http://www.tigr.org/softlab/glimmerm.htm), and mpeptideat (Mihaila Fertea and Steven Salzberg, contact mpeptideat@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to indicate the likelihood of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by RepeatMasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Location/Qualifiers		
FEATURES			
source			
	1..75948		
	/organism="Arabidopsis thaliana"		
	/cultivar="Colombia2"		
	/db_xref="taxon:3702"		
	/chromosome="1"		
	/map="m213"		
	/clone="F19K6"		
	1..1723		
gene			
	/gene="F19K6.1"		
mRNA			
	join(<1..97,223..405,517..723,811..958,1057..1187,1271..1360,1454..1534,1628..>1723)		
	/gene="F19K6.1"		
CDS			
	join(<1..97,223..405,517..723,811..958,1057..1187,1271..1360,1454..1534,1628..>1723)		


```

Db 42387 GGAGGTGAGGTGTTAAAGTCTGCTTCGGTGGTCCAAAGAGAGCGCTTTTTCATC 42446
QY 196 CACTGTGATGTTACTTAATGAAGACGGTGTAAANATGCGGTGACACACAGTTCATCACC 255
Db 42447 CATGGCGATGTTAGAGTGAAGATGACATTAGCAATGCGGTTGACTTTCAGTCAAAAAT 42506
QY 256 TATGGAACACTGGACATTTATGTTCCAGCAATGACAGGAATTTCTGATCCCAACAGGCCCGC 315
Db 42507 TTTGGACGCTTGATATACATTATACAAATGACAGGATATGAGGACCGCTGCCCTGAT 42566
QY 316 ATCATAGACACGAAAGACGCTTTGAACGCTTCTCAGGTAAATGTAAACCGGAGTT 375
Db 42567 ATTCGTAATTAATGTTTACGTGATTCGAGATCACCTTTGATGTGAATGTGAAAGAGCT 42626
QY 376 TTCCTATGATGAAGACGACGACGCTGTATGATTCAGACGACGAGTGGCAACATAAT 435
Db 42627 TTTCTAAGCATGAACATGACGCTCGTAAAGATACCGGAGAGAAAGAGGTCGATAGT 42686
QY 436 TCCACTGCTAGTTTAAGCTCAACTATGGGTGGTGGTCTTCACATGCTTATGTGTTCA 495
Db 42687 TCCTTATGATGTTGGAGGTGTTGGAGGCGTTGGTCCCAATCTTATGTTGTTCC 42746
QY 496 AAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCCGAGTCCGACAAATTTGGCATT 555
Db 42747 AAGCATGCTGTGTAGGCTTACTAGGAGTGTTCGACGCGAGCTTGGACAGCACGGGATA 42806
QY 556 AGGTTAATGTTGTCTCTTCGCGCTTCCTAC-----GGCTTTAGGCAAGAAATTT 609
Db 42807 CGTGTGAAGTGTGTTTCGCTTACGCGTTGCAACTAACTCCCTTGGCTCATTTGCCG 42866
QY 610 TCAGGATTAATAAATGAAGAAGATTTGAGAATGTAATAAATTTGC-----GGGA 660
Db 42867 GAGGAAGAAGAACGAGGATGCATTTGTTGTTTCAGGAATTTGCTCTGCAACCGC 42926
QY 661 AATTTGAAGTCCAAATTTAATGTTGAGGATGTTCCCAATGACGCTCTTATCTGCT 720
Db 42927 AATCTAAGGGGGTGGAACTGACGGTGTGATGATGATGATGATGATGATGATGATGATGAT 42986
QY 721 AGTGATGAGGCAAAATACGTTAGTGGACACAAATCTGTTCAATGATGAGGCTTCAGCGTC 780
Db 42987 AGCGATGACTCGCGTACATAAGCGAGATAATTTGATGATGATGATGATGATGATGATGAT 43046
QY 781 TGCAATTCGTAATCAAGTGTTCCAAT 808
Db 43047 ACTAACCACCTCTTAAGTCTTCAGAT 43074

```

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	588.5	42.2	337	1	US-08-440-856A-3	Sequence 3, Appli
2	582.5	41.8	333	1	US-08-440-856A-4	Sequence 4, Appli
3	500	35.9	359	1	US-08-440-856A-8	Sequence 8, Appli
4	369	26.5	256	4	US-09-504-358-14	Sequence 14, Appl
5	369	26.5	256	4	US-09-504-314-14	Sequence 14, Appl
6	349.5	25.1	262	4	US-09-363-1898-6	Sequence 6, Appli
7	341.5	24.5	261	4	US-09-468-738A-29	Sequence 29, Appli
8	341.5	24.5	261	4	US-09-940-019-29	Sequence 29, Appli
9	331.5	23.8	267	4	US-09-134-001C-5042	Sequence 5042, Ap
10	324.5	23.3	257	4	US-09-134-001C-3562	Sequence 3562, Ap
11	320	23.0	255	4	US-08-815-225-4	Sequence 4, Appli
12	317.5	22.8	274	4	US-09-134-001C-4431	Sequence 4431, Ap
13	315	22.6	271	2	US-07-637-865-2	Sequence 2, Appli
14	312.5	22.4	244	1	US-08-375-962B-13	Sequence 13, Appl
15	312.5	22.4	244	2	US-08-562-114B-13	Sequence 13, Appl
16	312.5	22.4	244	4	US-08-729-594A-13	Sequence 13, Appl
17	312.5	22.4	244	4	US-08-937-993-13	Sequence 13, Appl
18	304.5	21.9	249	4	US-09-134-001C-4825	Sequence 4825, Ap
19	301.5	21.6	246	3	US-09-238-810-2	Sequence 2, Appli
20	301.5	21.6	246	4	US-09-572-810A-2	Sequence 2, Appli
21	294.5	21.1	252	3	US-08-822-322-8	Sequence 8, Appli
22	294.5	21.1	252	4	US-09-466-109-8	Sequence 8, Appli
23	293	21.0	256	1	US-08-594-808B-7	Sequence 7, Appli
24	272	19.5	283	4	US-09-367-012-1	Sequence 1, Appli
25	272	19.5	283	4	US-09-777-157A-1	Sequence 1, Appli
26	265.5	19.1	243	4	US-09-239-052-2	Sequence 2, Appli
27	265.5	19.1	292	4	US-09-468-738A-2	Sequence 2, Appli

```
Db 106 VRCDSVEDDVRRAVDWALSRHGRDLVYCNAGVLGRQTRAARSILSFDAAEFRDLRV 165
QY 121 NVTGVFLCMKHAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAVLALTRNLAVEL 180
Db 166 NALGAALCMKHAARAPRAGSIVSVASVAAVLGGGLPHAYTASKHAIVGLTKNAACEL 225
QY 181 GQFGRVNCISPFGLPTA-----LGKKFSGIKNEEFENVI 216
Db 226 RAHGVVNCVSPFGVATPMLINAWROGHDDATADADRLDLDTVPDQDEVEKMEEVV 285
QY 217 NFAGNLKCPKPNVEDVANAALYLASDEAKYVSGHNLFDGGFSVCNSVI 265
Db 286 RGLATLKGPTLRPDIAEAFLASDEARIISGHNLVVDGGVTTSRNLI 334

RESULT 2
US-08-440-856A-4
; Sequence 4, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE. N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/440,856A
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLMAN, ROBERT A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 05463-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1517
; TELEFAX: (202) 887-0763
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-440-856A-4

Query Match 41.8%; Score 582.5; DB 1; Length 333;
Best Local Similarity 42.2%; Pred. No. 1.le-54;
Matches 121; Conservative 50; Mismatches 89; Indels 27; Gaps 4;

QY 5 TAFARRLEGKVALITGASGIGETTAKLFSQHGAKVAIADVDQLGHSVVEAIGTSNSTY 64
Db 46 TPKPKRLGKVAIVTGGARGIGEAIVRLFKHGAKVVIADIDDAAGEALAAALG- PHVGF 104
QY 65 IHCDDVTNEDGVKNAVDNVTSTYTKLIDIMFSNAGI---SDPNRPRIIDNEKADFVRLSVN 121
Db 105 VRCDSVEEDVERAVERAVARYGRDLVLCNNAGVLGRQTRAARKSILSPDAGEFDRVLVN 164
QY 122 VTGVFLCMKHAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAVLALTRNLAVELG 181
Db 165 ALGAALCMKHAALAMTORRAGSIISVASVAGVLGGGLPHAYTASKHAIVGLTKNAACELG 224
```

```
QY 182 QGIRVNCISPFGLPTAL-----GKKFS-----GIKNEEFENVINFE 218
Db 225 AHGIRVNCISPFGVATPMLINAWROGHDAADADADLDLIAVPSDQDEVEKMEEVVRG 284
QY 219 AGNLKGPKNVEDVANAALYLASDEAKYVSGHNLFDGGFSVCNSVI 265
Db 285 LATLKATGLRPDIAEAFLASDDSRYSISGHNLVVDGGVTTSRNLI 331

RESULT 3
US-08-440-856A-8
; Sequence 8, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE. N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/440,856A
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLMAN, ROBERT A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 05463-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1517
; TELEFAX: (202) 887-0763
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-440-856A-8

Query Match 35.9%; Score 500; DB 1; Length 359;
Best Local Similarity 38.8%; Pred. No. 1.le-45;
Matches 112; Conservative 40; Mismatches 107; Indels 30; Gaps 5;

QY 5 TAFARRLEGKVALITGASGIGETTAKLFSQHGAKVAIADVDQLGHSVVEAIGTSNSTY 64
Db 50 TPKPKRLGKVAIVTGGARGIGEAIVRLFKHGAKVVIADIDDAAGEALAAALGPK-VXF 108
QY 65 IHCDDVTNEDGVKNAVDNVTSTYTKLIDIMFSNAGI---SDPNRPRIIDNEKADFVRLSV 120
Db 109 VRCDSVEXDVXRAVAXXXRXGRDLVXCNNAGVLGRQTRAAXSILSFDAXEFDRLRV 168
QY 121 NVTGVFLCMKHAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAVLALTRNLAVEL 180
Db 169 NALGAALCMKHAAXAMXXRRAGSIXSVASVAVLGGGLPHAYTASKHAIVGLTKNAACEL 228
QY 181 GQFGRVNCISPFGLPTA-----LGKKFSGIKNEEFENVI 216
Db 229 XH-GXRVNCXSPFGVATPMLINAWROGHDXXTADXXDXDLDDXXVPSDQDEVEKMEEVV 287
QY 217 NFAGNLKCPKPNVEDVANAALYLASDEAKYVSGHNLFDGGFSVCNSVI 265
```



```
; Patent No. 6312933
; GENERAL INFORMATION:
; APPLICANT: Kimoto, No. 6312933ihiro
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Mitsuhashi, Kazuya
; TITLE OF INVENTION: NOVEL CARBONYL REDUCTASE, METHOD FOR PRODUCING SAID ENZYME, DNA
; TITLE OF INVENTION: ENCODING SAID ENZYME, AND METHOD FOR PRODUCING ALCOHOL USING SAID
; TITLE OF INVENTION: ENZYME
; FILE REFERENCE: 06501-050001
; CURRENT APPLICATION NUMBER: US/09/468,738A
; PRIOR FILING DATE: 1999-12-21
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: JP 1998-171160
; PRIOR APPLICATION NUMBER: JP 1998-363130
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0, reformatted using WordPerfect 5.1
; SEQ ID NO 29
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; ORGANISM: Bacillus subtilis
US-09-468-738A-29

Query Match 24.5%; Score 341.5; DB 4; Length 261;
Best Local Similarity 34.4%; Pred. No. 8.8e-29;
Matches 93; Conservative 50; Mismatches 98; Indels 29; Gaps 10;

QY 11 LEGKVALITGASGIGETTAKLFSOHGAKVAI---ADVOD--ELGHSVVEAIGTSNSTYI 65
DB 5 LKGKVVAITGAASGLGKAMAREGKEQAKVINYYSNKQDPNEVKEVIKAGG--EAVVV 62
QY 66 HCDVTNEDGVKNVADNTVSTYGLDIMEFNAGISDPNRPRIIDNEK--ADFERVLSVNV 123
DB 63 QGDVTKEDVKNIVQTAKEFGFLDINNAGLENP----VPSHEMPLKDWKVICTNLT 118
QY 124 GVFLCMKHAARVMIPAR-SGNIISTASLSTMGSGSHAYCGSKHVALTRNLAVELGQ 182
DB 119 GAFGSRRAIKFYENDIKGNVIMSSVHEVIPWPLFVHYAASKGIGKILMTETLALAYAP 178
QY 183 FGIRVNCLSPLGLPTAL-GKKFGSGIKNEEFENVINFAGNLKGPKNFVEDVANAALYLAS 241
DB 179 KGI RVNNGPGAINTPINAEKFPADPKADVESMIPM-GYIGEP----EETAAVAAMLAS 233
QY 242 DEAKYVSGHNLFDGGSFVCSNVIKVFQYP 271
DB 234 KEASYVTGTLFADGGMT-----QYP 254

RESULT 9
US-09-134-001C-5042
; Sequence 5042, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5042
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5042

Query Match 23.8%; Score 331.5; DB 4; Length 267;
Best Local Similarity 32.1%; Pred. No. 1.1e-27;
Matches 88; Conservative 55; Mismatches 96; Indels 35; Gaps 9;

QY 11 LEGKVALITGASGIGETTAKLFSOHGAKVAIADVDDELGHSVVEAIGTSNSTY----- 64
DB 9 LENKVVLTGAATGIGAKIAENFGKAKVVI-NYRSDRHSEIEEIKQTAKFGGQTLA 67
QY 65 IHCVDVTNEDGVKNVADNTVSTYGLDIMEFNAGI---SDPNRPRIIDNEKADFERVLSVNV 122
DB 68 VOGDVSEEDIKRMETIINHFTGLDIIINNAGFENSIPHEMSID----DWQKVIDINL 123
QY 123 TGVFLCMKHAARVMIPAR-RSGNIISTASLSTMGSGSHAYCGSKHVALTRNLAVELG 181
DB 124 TCAIVGSRRETINQFLKENKGTIINISVSHDTPWPNVHYAASKGGLKLMETMSMEYA 183
QY 182 QFGRVNCLSPLGLPTALGK-KFSGKIKNEEFENVINFAGNLKGPK---FNVEDVANAAL 237
DB 184 QYGRINNISPGAIVTEHTKEFSDPTRETIKMI-----PAREIGNAQDVANAVL 235
QY 238 YLASDEAKYVSGHNLFDGGSFVCSNVIKVFQYP 271
DB 236 FLSSDLASYIHGTLTYVDGG-----MMNYP 260
```



```

Db 64 DISDDQVQKOFSEKIAQEFHGVDFVFNAGV-DNAGAGRIHEYVPEVFDKIMAVDMRGTEL 122
QY 128 CMKHAARVMPARSNIISTASLSSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGIRV 187
Db 123 VTKFLPLMM-KQGSINTSFQSOADLYRSYNAAGGVNFTKSTIAIEYGRINIRA 181
QY 188 NCLSPGLPTALGKFKSGIKNEE-----EFENVINFAGNLKPKFKNVEDVANAALYLAS 241
Db 182 NATAPGTIETPLVDNLASTSDEEAGQTFRENQKWVTLPLRLGTP-----DEVGKLVAFLAS 237
QY 242 DEAKYVSGHNLFDGG 257
Db 238 DSSFTITGIRIDGG 253

```

```

RESULT 13
US-07-637-865-2
; Sequence 2, Application US/07637865
; Patent No. 5942427
; GENERAL INFORMATION:
; APPLICANT: Otake, Hideko
; APPLICANT: Koyama, Yasuji
; APPLICANT: Horiuchi, Tatsuo
; APPLICANT: Nakano, Eiichi
; TITLE OF INVENTION: N-Acetylmannosamine Dehydrogenase Gene and
; TITLE OF INVENTION: No. 5942427el Recombinant DNA as well as a Method for
; TITLE OF INVENTION: Product of N-Acetylmannosamine Dehydrogenase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: One Thomas Circle, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/637,865
; FILING DATE: 1990/12/27
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 01-338267
; FILING DATE: 28 DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Skappert, Joseph M.
; REGISTRATION NUMBER: 29,864
; REFERENCE/DOCKET NUMBER: 0118.033382
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-296-7300
; TELEFAX: 202-296-7830
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 271 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-637-865-2

```

```

Query Match 22.6%; Score 315; DB 2; Length 271;
Best Local Similarity 32.4%; Pred. No. 6.9e-26;
Matches 83; Conservative 46; Mismatches 113; Indels 14; Gaps 5;
QY 10 RLECKVALITGASGIGETTAKLFQSHGAKVAIADVQDELGHVSVEAIGTSNYSYIHCVDV 69
Db 12 FLACKAIVTGAAGGIGRATVEAYLEGAASVAMDAPLAATRYEPPG---AIPACDL 68
QY 70 TNEGVKNAVDNTVSYTGKIDIMFSNAGISDPNRPRIIDNEKADFERVLSVNTGVFLCM 129

```

```

Db 69 ADRAAIDAAMADAVARLGLDILVAGGALKG-GTGNFLDSDADWDRYVDVNMGTGLTC 127
QY 130 KHAARYWIFA-----RSGNIISTASLSSTMTGGSSHAYCGSKHAVLALTRNLAVELGQF 183
Db 128 RAGARAWAAGAGKOGKSARIIITIGSVNSFMAEPEAAAYAAKGVAMLTRAMAVDLARH 187
QY 184 GIRVNCLSFPGPLTALGKFKNEEEFENVINFAGNLKPKFKNVEDVANAALYLASDE 243
Db 188 GILVNMIAFG--PVDVTGNNTGYSERLAEQVLDEVA--LGRPGLPEEVATAAFLAEDG 243
QY 244 AKYVSGHNLFDGGFS 259
Db 244 SSPITGSTITIDGGLS 259

```

```

RESULT 14
US-08-375-962B-13
; Sequence 13, Application US/08375962B
; Patent No. 5731195
; GENERAL INFORMATION:
; APPLICANT: SIMON, ANDRAS; HELLMAN, ULF; WERNSTEDT,
; APPLICANT: CHRISTER, ERIKSSON, ULF
; TITLE OF INVENTION: Isolated Nucleic Acid Molecule
; TITLE OF INVENTION: Which Codes for A 32 kDa Protein Having 11-CIS Retinol
; TITLE OF INVENTION: Dehydrogenase Activity, and Which Associates With P85,
; TITLE OF INVENTION: a Portion of a Retinol Binding Protein Receptor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect (ASCII standard)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,962B
; FILING DATE: 20-January-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/258,418
; FILING DATE: 6-October-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: LUD 5372.1 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 688-3884
; INFORMATION FOR SEQ ID NO: 13:
; LENGTH: 244 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: E.Coli 3-oxoacyl[acyl carrier protein]
; NAME/KEY: reductase (FABG)
US-08-375-962B-13

```

```

Query Match 22.4%; Score 312.5; DB 1; Length 244;
Best Local Similarity 31.6%; Pred. NO. 1.1e-25;
Matches 78; Conservative 51; Mismatches 107; Indels 11; Gaps 5;
QY 12 EGKVALITGASGIGETTAKLFQSHGAKVAIADVQDELGHVSVEAIGTSNYSYIHCVDV 71
Db 4 EGKIALVTGASRGIGRAIAETLAARGKV-IGTATSENGAQAISDYLGAANGKGLMNVTD 62
QY 72 EDGVKNAVDNTVSYTGKIDIMFSNAGISDPN-RPRIIDNEKADFERVLSVNTGVFLCMK 130

```


Db 63 PASIESVLEKIRAEFGVDILVNNAGITRDNLMLRMKDEWND---IETNLSSVFLSK 119
QY 131 HAARVMIPARSGNIISTASLSSTMGSSSHAYCGSKHAVLALTRNLAVELGQFGIRVNCL 190
Db 120 AVNRAMMKRHRGRIITIGSVVGTMGNGQANYAAAKAGLIGFSKSLAREVASRGITVNVV 179
QY 191 SPFGPLTALGKFKSGIKNEEEFENVINFGNLKGPKNVEDVANAALYLASDEAKYVSGH 250
Db 180 APGFIETDMTRALSDQDQAGILAQVP--AGRLGG---AQEIANAVAFSLASDEAAVITGE 233
QY 251 NLFIDGG 257
Db 234 TLHVNGG.240

RESULT 15

US-08-562-114B-13
; Sequence 13, Application US/08562114B
; Patent No. 5972646
; GENERAL INFORMATION:
; APPLICANT: ERIKSSON ET AL.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE WHICH CODES FOR A
; TITLE OF INVENTION: 32 KDA PROTEIN HAVING 11-CIS RETINOL HYDROGENASE
; TITLE OF INVENTION: ACTIVITY, AND WHICH ASSOCIATES WITH P63, A
; TITLE OF INVENTION: PORTION OF A RETINOL BINDING PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect 5.1 and ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/562,114B
; FILING DATE: 22-No. 5972646ember-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/375,962
; FILING DATE: 20-January-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohli, Vineet
; REGISTRATION NUMBER: 37,003
; REFERENCE/DOCKET NUMBER: LUD 5372.2 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: E.coli 3-oxoacyl[acyl carrier protein]reductase
; NAME/KEY:
US-08-562-114B-13

Query Match 22.4%; Score 312.5; DB 2; Length 244;
Best Local Similarity 31.6%; Pred. NO. 1.1e-25;
Matches 78; Conservative 51; Mismatches 107; Indels 11; Gaps 5;

QY 12 EGKVALITGASGIGETTAKLFSQHGAKVAIAVDQDELGHVSVEAIGTSGNSTVIHCDVTN 71
Db 4 EGKIALVTGASRGIGRAIAETLAARGKV-IGTATSENGAQAISDYLGANGKGLMLNVTD 62
QY 72 EDGVKNVAVONTVSTYKGLDIFMSNAGISDPN-RPRIIDNEKADFERVLSVNVYTGFLCMK 130

Db 63 PASIESVLEKIRAEFGVDILVNNAGITRDNLMLRMKDEWND---IETNLSSVFLSK 119
QY 131 HAARVMIPARSGNIISTASLSSTMGSSSHAYCGSKHAVLALTRNLAVELGQFGIRVNCL 190
Db 120 AVNRAMMKRHRGRIITIGSVVGTMGNGQANYAAAKAGLIGFSKSLAREVASRGITVNVV 179
QY 191 SPFGPLTALGKFKSGIKNEEEFENVINFGNLKGPKNVEDVANAALYLASDEAKYVSGH 250
Db 180 APGFIETDMTRALSDQDQAGILAQVP--AGRLGG---AQEIANAVAFSLASDEAAVITGE 233
QY 251 NLFIDGG 257
Db 234 TLHVNGG 240

Search completed: February 14, 2003, 09:21:39
Job time : 17 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:54 ; Search time 33 Seconds
(without alignments)
1704.572 Million cell updates/sec

Title: US-09-673-918A-2

Perfect score: 1393
Sequence: 1 MQLTAFARLEKGVKALITG.....IDGFSVNSVIKVFQVPS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertibrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1364	97.9	277	Q94KL7	Q94kl7 forsythia i
2	734	52.7	267	P93697	P93697 vigna ungui
3	700.5	50.3	271	Q94G09	Q94g09 cucumis sat
4	693.5	49.8	271	Q94G10	Q94g10 cucumis sat
5	685	49.2	284	O50038	O50038 nicotiana t
6	677	48.6	284	O82465	O82465 ipomoea tri
7	673	48.3	277	Q9SBD8	Q9sbd8 ipomoea tri
8	670	48.1	284	Q9SBD6	Q9sbd6 ipomoea tri
9	668	48.0	277	Q9SBD7	Q9sbd7 ipomoea tri
10	665.5	47.8	273	Q9SBD0	Q9sbd0 ipomoea tri
11	658.5	47.3	285	Q9SBD0	Q9sbd0 citrullus l
12	652.5	46.8	278	Q9CK86	Q9ck86 arabidopsis
13	625.5	44.9	303	Q9SCU0	Q94kl8 podophyllum
14	597.5	42.9	283	Q9ZRL7	Q9scu0 arabidopsis
15	593.5	42.6	259	Q93Y47	Q9zrl7 arabidopsis
16	590	42.4	300	Q9LW35	Q93y47 digitalis l
					Q9lw35 arabidopsis

17	590	42.4	331	10	P93795	P93795 tripsacum d
18	579.5	41.6	259	10	Q9LEG3	Q9leg3 lycopersico
19	575.5	41.3	296	10	Q9LS70	Q9ls70 arabidopsis
20	574.5	41.2	257	10	Q94K41	Q94k41 arabidopsis
21	567	40.7	257	10	Q80713	Q80713 arabidopsis
22	565	40.6	306	10	Q9LW34	Q9lw34 arabidopsis
23	560	40.2	306	10	Q94L54	Q94le4 arabidopsis
24	559	40.1	251	10	Q40133	Q40133 lycopersico
25	549.5	39.4	258	10	Q80714	Q80714 arabidopsis
26	543.5	39.0	268	10	Q9SQF9	Q9sqf9 pisum sativ
27	542.5	38.9	268	10	Q9SQJ3	Q9sqj3 pisum sativ
28	534.5	38.4	294	10	Q94FD0	Q94fd0 zea mays su
29	534.5	38.4	294	10	Q94FC9	Q94fc9 zea mays su
30	534.5	38.4	294	10	Q93W09	Q93w09 zea mays su
31	534.5	38.4	294	10	Q93VX5	Q93vx5 zea mays su
32	534.5	38.4	294	10	Q93VX4	Q93vx4 zea mays su
33	533.5	38.3	296	10	Q94FC8	Q94fc8 zea mays su
34	533.5	38.3	296	10	Q94FC7	Q94fc7 zea mays su
35	533.5	38.3	296	10	Q93W76	Q93w76 zea mays su
36	533.5	38.3	296	10	Q93W75	Q93w75 zea mays su
37	508.5	36.5	281	10	Q41345	Q41345 silene prat
38	503.5	36.1	281	10	Q42407	Q42407 silene lati
39	499.5	35.9	234	10	Q40590	Q40590 nicotiana t
40	490.5	35.2	272	10	Q9M1K9	Q9mlk9 arabidopsis
41	487.5	35.0	272	10	O24452	O24452 arabidopsis
42	471.5	33.8	256	10	Q9SQJ2	Q9sqj2 pisum sativ
43	468.5	33.6	329	10	P93796	P93796 tripsacum d
44	432	31.0	253	16	Q91376	Q91376 pseudomonas
45	431	30.9	160	10	Q8S9B7	Q8s9b7 ipomoea tri

ALIGNMENTS

RESULT 1

Q94KL7 ID Q94KL7 PRELIMINARY; PRT; 277 AA.

AC Q94KL7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Stem secolariciresinol dehydrogenase (Fragment).
OS Forsythia intermedia (Border forsythia).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Oleaceae; Forsythia.
OX NCBI_TaxID=55183;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21201084; PubMed=11278426;
RA Xia Z.O., Costa M.A., Pelissier H.C., Davin L.B., Lewis N.G.;
RT "Secoisolariciresinol Dehydrogenase Purification, Cloning, and
Functional Expression. IMPLICATIONS FOR HUMAN HEALTH PROTECTION.";
RL J. Biol. Chem. 276:12614-12623(2001).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
DR EMBL: AF352735; AAK38665.1; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
FT NON_TER 277
SQ SEQUENCE 277 AA: 29256 MW; 98885C210CAFE2EB CRC64;

Query Match 97.9%; Score 1364; DB 10; Length 277;
Best Local Similarity 98.2%; Pred. No. 2.9e-95;
Matches 267; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 QURTAFAARLEKGVKALITGASIGETTAKLFSOHGAKVAIADVDQELGHSVVEAIGTSN 61

Db 6 QVLTATARLEKGVKALITGASIGETTAKLFSOHGAKVAIADVDQELGHSVVEAIGTSN 65

Qy 62 STYIHCDVTNEDGVKNAVDNTVSTYTKGLDINFNSAGISDPNRPRIIDNEKADFERYLSVN 121

```

Db 66 STYIHCVDVTNEDGVKNADVNTVSTYKLDIMFNSAGISDPNRPRIIDNEKADFERVESFN 125
QY 122 VTGVFLCMKHAARVMIPARSGNIISTASLSSTMGSSSHAYCGSKHAYLALTRNLAVELG 181
Db 126 VTGVFLCMKHAARVMIPARSGNIISTASLSSTMGSSSHAYCGSKHAYLALTRNLAVELG 185
QY 182 QFGRVNCVLPFGALPGALGKFKSGIKNEEFENVINFGAGNLKGPKNFVEDVANALYLAS 241
Db 186 QFGRVNCVLPFGALPGALGKFKSGIKNEEFENVINFGAGNLKGPKNFVEDVANALYLAS 245
QY 242 DEAKYVSGHNLFDGGFSGVNSVIKVFQYFDPDS 273
Db 246 DEAKYVSGHNLFDGGFSGVNSVIKVFQYFDPDS 277

RESULT 2
P93697 ID P93697 PRELIMINARY; PRT; 267 AA.
AC P93697;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE CPRD12 protein.
OS Vigna unguiculata (Cowpea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX NCBI_TaxID=3917;
RN [1]
RA luchi S., Yamaguchi-Shinozaki K., Urao T., Shinozaki K.;
RT "Characterization of two cDNAs for novel drought-inducible genes in
RL J. Plant Res. 109:415-424 (1996).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DB EMBL; D88121; BAA13541.1; -.
DR HSSP; P19992; LHDC.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 267 AA; 28484 MW; 8C4A48A2E9F41B64 CRC64;

Query Match 52.7%; Score 734; DB 10; Length 267;
Best Local Similarity 56.3%; Pred. No. 1e-47;
Matches 147; Conservative 41; Mismatches 67; Indels 6; Gaps 2;

QY 9 RLREGKVALITGASGIGETAKLFSQHGAKVAIADVQDELGSHVVEAIGTSYTHCD 68
Db 12 KRLGKVAIITGGASGIGETAKLFSQHGAKVAIADVQDELGSHVVEAIGTSYTHCD 69
QY 69 VTNEDGVKNADVNTVSTYKLDIMFNSAGISDPNRPRIIDNEKADFERVLSVNTGVFLC 128
Db 70 VTKEDIEKVDVTSKFGKLDIMFNAGTGDFFKSLDNTKSDFERVLSVNLVGPFLG 129
QY 129 MKHAARVMIPARSGNIISTASLSSTMGSSSHAYCGSKHAYLALTRNLAVELGQFGRVN 188
Db 130 TKHAARVMIPARSGNIISTASVAGCIGGATHAYTSSKHALVGLTKTAVELGQFGRVN 189
QY 189 CLSPFGLPTALGKFKSGIKNEEFENVINFGAGNLKGPKNFVEDVANALYLASDEAKYVS 248
Db 190 CVSPFAIVTPLNKNLDEEGVRKTYMNLKGYFVPV----NDVAEALYLASDESKEVS 245
QY 249 GHNLFIDGGFSGVNSVIKVFQ 269
Db 246 SHNLVDGGLINSVGFPMFE 266

RESULT 3
Q94G09 ID Q94G09 PRELIMINARY; PRT; 271 AA.
AC Q94G09;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Sex determination protein..
OS Cucumis sativus (Cucumber).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=3659;
RN [1]
RA Sun J.-Q., Li Q.-Z., Li X.-G.;
RT "Gene cloning and expression of CSG in cucumber.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DB EMBL; AF286651; AAK83036.1; -.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 271 AA; 28997 MW; 7CD6ED7F52AD6224 CRC64;

Query Match 50.3%; Score 700.5; DB 10; Length 271;
Best Local Similarity 53.7%; Pred. No. 3.6e-45;
Matches 146; Conservative 44; Mismatches 75; Indels 7; Gaps 4;

QY 1 MOLRTAFARLEKGVVALITGASGIGETAKLFSQHGAKVAIADVQDELGSHVVEAIGTS 60
Db 3 IQLLPALARLEKGVAVITGARGIGEQTAKLFFKGAHVVIADIQDLGTLCKDLQGS 62
QY 61 NSTYIHCVDVTNEDGVKNADVNTVSTYKLDIMFNSAGI--SDPNRPRIIDNEKADFERVL 118
Db 63 SSVFVHCDVTKEKDVETAVDTAVTSKYKLDIMLNAGVFEESPFD-FLKDDPLTFQVW 121
QY 119 SVNVTGVFLCMKHAARVMIPARSGNIISTASLSSTMGSSSHAYCGSKHAYLALTRNLAV 178
Db 122 NVNLVGAFLGTGKHAARVMKPAGRSVITVTASICVIGGIGTTHAYTSSKHVGLMLRNAV 181
QY 179 ELGQFGRVNCVLPFGALPGALGKFKSGIKNEEFENVINFGAGNLK-GPKFNVEDVANAL 237
Db 182 DLGRYGRVNCVSPNVVPTGMRKLFKVKDGGEP---SFYWSLKNGLDILREEDYGEAVV 238
QY 238 YLASDEAKYVSGHNLFDGGFSGVNSVIKVFQ 269
Db 239 YLGDESKVSGNLIVDGGFTVYNQALCSFR 270

RESULT 4
Q94G10 ID Q94G10 PRELIMINARY; PRT; 271 AA.
AC Q94G10;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE CTA.
OS Cucumis sativus (Cucumber).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=3659;
RN [1]
RA Sun J.-Q., Li Q.-Z., Li X.-G., Zhang X.-S.;
RT "Gene cloning and expression of CTA in cucumber.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DB EMBL; AF286650; AAK83035.1; -.

```


Qy 248 SGHNLFDGGSVCNSVIK 267
 Db 251 SGLNLLIDGGFTTTNIAFQV 270

RESULT 7
 Q9SBD8 PRELIMINARY; PRT; 277 AA.

AC Q9SBD8: 2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Short-chain alcohol dehydrogenase.
 GN SSP.
 OS Ipomoea trifida.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
 OX NCBI_TaxID=35884;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Koyama Y., Kadota N., Sassa H., Kakeda K., Tsuchiya T., Ishimizu T.,
 RA Kondo K., Norioka S.;
 RT "S-luciferin specific stigma protein (SSP) from sporophytic self-
 RT incompatibility plant, Ipomoea trifida, is a member of short-chain
 RT alcohol dehydrogenase family".
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 DR EMBL; AF072448; AAC35341.1; -.
 DR HSP; P19992; LHDC.
 DR InterPro; IPR002198; ADH_short.
 DR Pfam; PF00106; adh_short; 1.
 DR PRINTS; PR00080; SDRFAMILY.
 KW Oxidoreductase.
 SQ SEQUENCE 277 AA; 29312 MW; 7A75BFE12A73D9AE CRC64;

Query Match 48.3%; Score 673; DB 10; Length 277;
 Best Local Similarity 51.7%; Pred. No. 4.5e-43;
 Matches 134; Conservative 42; Mismatches 83; Indels 0; Gaps 0;

Qy 9 RLREGKVALITGASGIGETTAKLFSQHGAKVAIADVDDELGHGSVVEAIGTSNTYIHCD 68
 Db 12 KRLGKVAIITGAANGIGATTARLFAHQGCKVLIADIDKNGHSAVEEIGPEYALFHCD 71
 Qy 69 VTNEGDKNAVNTVSYGKLDIMFNSAGISDPNRPRIIDNEKADFVLSVNVTVGVFLC 128
 Db 72 VRIESHVQHADVTVSYGKLDIMFNSAGVAGSRDTSILEASPENINLVFETNVFGAFC 131
 Qy 129 MKHAARVWIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGRVN 188
 Db 132 AKHAARVWIPARKGVIFSAASEVFGITSDTYTASKCAVVGKSLCVMKGYGKIKAN 191
 Qy 189 CLSPFGLPTALGKFKSGIKNEEFENVFNAGNLGPKFNVEDVANAALYLASDEAKYVS 248
 Db 192 CVSPYVILTGLGMSIMPTQDRKLAEEIVAEASNFKGKLTITTEDVAEALYLAGDESKFVS 251
 Qy 249 GHNLFDGGSVCNSVIK 267
 Db 252 GLNLLIDGGFTTTNIAFQV 270

RESULT 8
 Q9SBD6 PRELIMINARY; PRT; 284 AA.

AC Q9SBD6: 2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Short-chain alcohol dehydrogenase.
 GN SSP.
 OS Ipomoea trifida.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
 OX NCBI_TaxID=35884;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Koyama Y., Kadota N., Sassa H., Kakeda K., Tsuchiya T., Ishimizu T.,
 RA Kondo K., Norioka S.;
 RT "S-luciferin specific stigma protein (SSP) from sporophytic self-
 RT incompatibility plant, Ipomoea trifida, is a member of short-chain
 RT alcohol dehydrogenase family".
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 DR EMBL; AF072450; AAC35343.1; -.
 DR HSP; P19992; LHDC.
 DR InterPro; IPR002198; ADH_short.
 DR Pfam; PF00106; adh_short; 1.
 DR PRINTS; PR00080; SDRFAMILY.
 KW Oxidoreductase.
 SQ SEQUENCE 284 AA; 29898 MW; 2B54BFFBAD6BAE4A CRC64;

Query Match 48.1%; Score 670; DB 10; Length 284;
 Best Local Similarity 51.7%; Pred. No. 7.8e-43;
 Matches 134; Conservative 41; Mismatches 84; Indels 0; Gaps 0;

Qy 9 RLREGKVALITGASGIGETTAKLFSQHGAKVAIADVDDELGHGSVVEAIGTSNTYIHCD 68
 Db 12 KRLGKVAIITGAANGIGATTARLFAHQGCKVLIADIDKNGHSAVEEIGPEYALFHCD 71
 Qy 69 VTNEGDKNAVNTVSYGKLDIMFNSAGISDPNRPRIIDNEKADFVLSVNVTVGVFLC 128
 Db 72 VRIESHVQHADVTVSYGKLDIMFNSAGVAGSRDTSILEASPENINLVFETNVFGAFC 131
 Qy 129 MKHAARVWIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGRVN 188
 Db 132 AKHAARVWIPARKGVIFSAASEVFGITSDTYTASKCAVVGKSLCVMKGYGKIKAN 191
 Qy 189 CLSPFGLPTALGKFKSGIKNEEFENVFNAGNLGPKFNVEDVANAALYLASDEAKYVS 248
 Db 192 CVSPYVILTGLGMSIMPTQDRKLAEEIVAEASNFKGKLTITTEDVAEALYLAGDESKFVS 251
 Qy 249 GHNLFDGGSVCNSVIK 267
 Db 252 GLNLLIDGGFTTTNIAFQV 270

RESULT 9
 Q9SBD7 PRELIMINARY; PRT; 277 AA.

AC Q9SBD7: 2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Short-chain alcohol dehydrogenase (S-locus linked stigma
 DE protein).
 GN SSP OR SSP10.
 OS Ipomoea trifida.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
 OX NCBI_TaxID=35884;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Koyama Y., Kadota N., Sassa H., Kakeda K., Tsuchiya T., Ishimizu T.,
 RA Kondo K., Norioka S.;
 RT "S-luciferin specific stigma protein (SSP) from sporophytic self-
 RT incompatibility plant, Ipomoea trifida, is a member of short-chain
 RT alcohol dehydrogenase family".
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H77-2;
 RA Tsuchiya T., Suwabe K., Watase T., Kagaya Y., Koyama Y.;

RT "Characterization of promoter region of the SSP gene from Ipomoea
RL trifida regulates stigma-specific expression."
CC Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL: AF072449; AAC35342.1; -.
DR EMBL: AB070221; BAB86916.1; -.
DR HSSP: P19992; 1HDC.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR OXidoreductase.
KW OXidoreductase.
SQ SEQUENCE 277 AA; 29257 MW; A0CDB434B31C666D CRC64;

Query Match 48.0%; Score 668; DB 10; Length 277;
Best Local Similarity 51.7%; Pred. No. 1.1e-42;
Matches 134; Conservative 41; Mismatches 84; Indels 0; Gaps 0;

QY 9 RRLGKVALITGASGIGETTAKLFSQHAKVAIADVDQDELGHVVEAIGTNSSTYIHCD 68
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 12 KRLEGKVAITGAANGIGATTARLFQAQHGCKVLIADIDDKNGLSVAEIGPEYALFIHCD 71

QY 69 VTNEGVKNVADNTVSTYKLDIMFSNAGISDPNRRPRIIDNEKADFERVLSVNVTVGVFLC 128
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 72 VRIESDVQNAVDTTVSRYGKLDIMFSNAGVAGSRDTSILEASPENINLVFETNVFGAFFC 131

QY 129 MKHAARVMIPARSGNIISTASLSSTMTGGSSSHAYCGSKHVALTRNLAVELGQFGQIRVN 188
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 132 AKHAARVMIPARKGVIFESAASEVFGITSDTYTAKCAVVGCLSKSLCVEMKGYGIKAN 191

QY 189 CLSPFGLPTALGKFKSGIKNEEFENVINFAGNLGPKFNVEDVANALYLASDEAKYVS 248
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 192 CVSPYVILTKLGSIMPTQDRKLAEBIVAESNFKGKTLTTEDVABAALYLASDESKFVS 251

QY 249 GHNLFIDGFGFVCSNVIKV 267
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 252 GLNLLIDGGFTTTNTAFQV 270

RESULT 10
Q9SBM0 PRELIMINARY; PRT; 273 AA.
ID Q9SBM0
AC Q9SBM0
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Wts2L.
GN Citrullus lanatus (Watermelon) (Citrullus vulgaris).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucurbitaceae; Cucurbitaceae;
OX NCBI_TaxID=3654;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim J., Kang H., Jun S., Lee J., An G.;
RT "Cloning of three gibberellin-regulated cDNAs from watermelon during
RT early seed development: down-regulated one cDNA and up-regulated two
RT cDNAs."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL: AB018559; BAA89230.1; -.
DR HSSP: P19992; 1HDC.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR OXidoreductase.
KW OXidoreductase.
SQ SEQUENCE 273 AA; 28835 MW; 904A835DD52B6E2 CRC64;

Query Match 47.8%; Score 665.5; DB 10; Length 273;
Best Local Similarity 51.1%; Pred. No. 1.6e-42;
Matches 137; Conservative 48; Mismatches 74; Indels 9; Gaps 4;

QY 9 RRLGKVALITGASGIGETTAKLFSQHAKVAIADVDQDELGHVVEAIGTNSSTYIHCD 68
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 11 RRLDGKVAITGASGIGTSVAVRIPHENGAKVIAIDQDEIGQIADQLG-DDVSVYIHC 69

QY 69 VTNEGVKNVADNTVSTYKLDIMFSNAGISDPNRRPRIIDNEKADFERVLSVNVTVGVFLC 128
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 70 VSKEDDVSNLVDAAVHRHGKLDIMYSNAGVLDKRSFGILDVTKSDLDKVLGVNVMGAFWG 129

QY 129 MKHAARVMIPARSGNIISTASLSSTMTGGSSSHAYCGSKHVALTRNLAVELGQFGQIRVN 188
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 130 AKHAARVMIPERKNGCLFTSSATNTIAGLSSHPYAAKCAVGLVRLNLAELGQHGIRVN 189

QY 189 CLSPFGLPTALGKFKSGIKNEEFENVINFAGNLGPKFNVEDVANALYLASDEAK 245
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 190 CVAPFVATGI---AGSRDPQAEATMTVTTWANLGRVLRKADDIKAAALYLASDDAN 245

QY 246 YVSGHNLFDGFGFVCSN-SVIRVFQYDP 272
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 246 YVSGNLVVDGGYSVYVNTMLTKLMD 273

RESULT 11
Q9C826 PRELIMINARY; PRT; 285 AA.
ID Q9C826
AC Q9C826
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Short chain alcohol dehydrogenase, putative.
GN FlrK6.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL: AC037424; AAG51536.1; -.
DR HSSP: P29132; 1DFI.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW OXidoreductase.
SQ SEQUENCE 285 AA; 30272 MW; 8B555DAED7386DF4 CRC64;

Query Match 47.3%; Score 658.5; DB 10; Length 285;
Best Local Similarity 50.9%; Pred. No. 5.8e-42;
Matches 139; Conservative 43; Mismatches 78; Indels 13; Gaps 5;

```

QY 8 ARLEGKVALITGASGIGETAKLFSQHGAKVAIADVDLGHSHVVEAI--GTSNST-- 63
Db 15 SORLLGVALITGATGIGESIVRLFHKGAKVCIVLDQDLGGEVCKSLRGESKETAF 74

QY 64 YIHCDVTNEDGVKNVNDVTSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSVNT 123
Db 75 FIHGDVVEDDISNAVDFAVKNGFTLDILINNAGLCGAPCDIRNYSLSFEFTFDVNVK 134

QY 124 GVFLCMKHAARVMIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLALRNLAVALGQF 183
Db 135 GAPTSMKHAARVMIPKSGISVLSVCSGVGVGVGPHSVGSKHAVLGLTRSAVAELGQH 194

QY 184 GIRVNCSPGLTALGKFKFGKNEEFENVI----NFA---GNLKGPKFENVEDVANA 236
Db 195 GIRVNCSPVAVATKLA--LAHLPEERTEDAFVCFRFAAANANLKGVELTVDDVANAV 252

QY 237 LYLASDEAKYVSGHNLFIDGGFSVCNSVYKQF 269
Db 253 LFLASDDSRISGDNLMIDGGFTCTNHSFKRVP 285

RESULT 12
Q94KL8
ID Q94KL8 PRELIMINARY; PRT; 278 AA.
AC Q94KL8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Rhizome secoisolariciresinol dehydrogenase (Fragment).
OS Podophyllum peltatum (Mayapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
OC Berberidaceae; Podophyllum.
OX NCBI_TaxID=35933;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=21201084; PubMed=11278426;
RA Xia Z.Q., Costa M.A., Pellissier H.C., Davin L.B., Lewis N.G.;
RT "Secoisolariciresinol Dehydrogenase Purification, Cloning, and
RT Functional Expression. IMPLICATIONS FOR HUMAN HEALTH PROTECTION.";
RL J. Biol. Chem. 276:12614-12623(2001).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
DR EMBL; AF352734; AAK38664.1; -.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; nad_binding.
DR PRINTS; PR000205; NAD_short.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
FT NON_TER 278
SQ SEQUENCE 278 AA; 29253 MW; DB735A376E112375 CRC64;

Query Match 46.8%; Score 652.5; DB 10; Length 278;
Best Local Similarity 52.2%; Pred. No. 1.6e-41;
Matches 133; Conservative 43; Mismatches 76; Indels 3; Gaps 3;

QY 10 RLEGKVALITGASGIGETAKLFSQHGAKVAIADVDLGHSHVVEAIGTSN-STYIHC 68
Db 13 RLQDKVAITGGAGGIGETAKLFRYKAVVIAADIADHCKVCNNIGSPDIVSFVHCD 72

QY 69 VTNEDGVKNVNDVTSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSVNTGVFLC 128
Db 73 VTQEDVRNLVDVTTHAKHGLDIFMSNAGISDPNRPRIIDNEKADFERVLSVNTGVFLC 132

QY 129 MKHAARVMIPARSGNIISTASLSS--TMGGSSHAYCGSKHAVLALRNLAVALGQFIRV 187
Db 133 AKHAARVMIPAKKSGIVFTASISSETAGEGVSHVYATKHAVLGLTSLCTELGEYIRV 192

QY 188 NCLSPFLGPTALCKKTSKNEEFENVIINFAAGNLKGPKNVEDVANAALYASDEAKYV 247
Db 193 NCVSPYIVASPLLTDFGVG--DSSRVEELAHQAANLKGTLLEAEDVADAVAYLAGDESKYV 251

```

```

QY 248 SGHNLFTDGGFSVCN 262
Db 252 SGLNLVIDGGYTRTN 266

RESULT 13
Q9SCU0
ID Q9SCU0 PRELIMINARY; PRT; 303 AA.
AC Q9SCU0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Short-chain alcohol dehydrogenase-like protein.
GN T18N14.60.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Delsen M., Berger C., Cooke R., Greillet F., Laudie M., Mewes H.W.,
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
DR EMBL; ALI32968; CAB63154.1; -.
DR HSP; P19992; IHDC.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short.1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
RN [1]
RP SEQUENCE 303 AA; 32151 MW; DC1BD28DABE38DD6 CRC64;

Query Match 44.9%; Score 625.5; DB 10; Length 303;
Best Local Similarity 47.5%; Pred. No. 2e-39;
Matches 131; Conservative 52; Mismatches 78; Indels 15; Gaps 6;

QY 5 TAFARRLEGKVALITGASGIGETAKLFSQHGAKVAIADVDLGHSHVVEAIGTSNST- 63
Db 26 TLYPKRLEGKVAITGGAGHIGKATVNLFPARHGATVVIADVDNVAGSLAKSLSSHKTS 85

QY 64 ---YIHCDVTNEDGVKNVNDVTSTYKGLDIFMSNAGI--SDPNRPRIIDNEKADFERV 118
Db 86 MVAFISCDVSVEADENLVNVTVARYGRDLILFNAGVLGDKKHKSLDFDADEFDHVM 145

QY 119 SVNYTGVFLCMKHAARVMIP-ARSGNIISTASLSTMTGGSSHAYCGSKHAVLALRNLA 177
Db 146 RVNVRGVGLGKMGKARAMIKRGFKGCIISTASVAGVMGMGPHTASKHAIVGLTKNAA 205

QY 178 VELQOFGIRVNCISPLGLTAL----GKKEFG--TKNE--BEFENVINFAAGNLKGPKNV 229
Db 206 CELYKIRVNCISPLGFAVSMVLNARKTSGDVEDDDVEEMEEFVRSANLKGTELRA 265

QY 230 EDVANAALYASDEAKYVSGHNLFIDGGFSVCNSVI 265
Db 266 NDIAEAALYASDESKEYVNGHNLVVDGGVTTARNCV 301

RESULT 14
Q92R17
ID Q92R17 PRELIMINARY; PRT; 283 AA.
AC Q92R17;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Putative alcohol dehydrogenase.
GN F4C21.6 OR AT4G03140.

```



```
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Habermann K., de la Bastide M., Huang E.N., Gnoj L., Schutz K.,
RA Preston R., Calma C., Martienssen R., Parnell L.D., Dedhia N.,
RA McCombie W.R.;
RT "Arabidopsis thaliana BAC F4C21 from chromosome IV near 17 cm.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL; AC005275; AAD14442.1; -.
DR EMBL; AL161496; CAB7799.1; -.
DR HSSP; PI9992; 1HDC.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PRO0080; SDRFAMILY.
DR PROSITE; P500061; ADH_SHORT; UNKNOWN_1.
DR Oxidoreductase.
KW OXIDOREDUCTASE.
SQ SEQUENCE 283 AA; 29994 MW; 93B0A97CECC859BC CRC64;

Query Match 42.9%; Score 597.5; DB 10; Length 283;
Best Local Similarity 47.0%; Pred. No. 2.3e-37;
Matches 126; Conservative 45; Mismatches 86; Indels 11; Gaps 4;

QY 9 RLEGKVALITGGASGIGETTAKLFSQHGAKVAIADVQDELGHSHVVEAIGTSNSTYIHC 68
Db :|||||:|||||: || :|||||:|||||: || :|||||: || :|||
16 RLEGKVALITGGASGIGATAGKFTSHGAKVLIADIQPIGRETEQELGPS-CAFFPCD 74
QY 69 VTNEDGVKNVDMVTSTYKGLDIFMFSNAGISDPNRPRIIDNEKADFERVLSVNVTVGVFLC 128
Db ||| : ||||| : ||| : ||||| : ||| : ||| : ||| : |||
75 VTKEIDIANAVDPAVSLHTKLDIMYNNAGIPCKTPPSIVDLNLFKDKVINTVNRGV 134
QY 129 MKHAARVMIPARSGNIISTASLSSTMGSSSHAYCGSKHVALALTRNLAVELGQFGIRVN 188
Db :|||||:|||||: |||: |||: |||: |||: |||: |||: |||
135 IKHAARVMIPRNSGSIICAGSVTGMGGGLAQHTYSVSKSAVIGIVRSTASELCKHR 194
QY 189 CLSPFGLPTA-----LCKKFSGKKEEFENFVNFAGNLKPKFNFVEDVANAALYLASDE 243
Db :||| : | : : : : : : : : : : : : : : : : : : : : : : : :
195 CISPFALTTSFVNDMRQIYPGV-DSRLTIQVQSTGVNLNGEVCETPDVANAAYLASDD 253
QY 244 AKTVSGHNLFDGFGFSCVNSVIKVFQYP 271
Db :||||:||||: ||| : ||| : ||| : ||| : ||| : ||| : |||
254 SKVYNGHNLVVDGGF----TTVKTLDFP 277

RESULT 15
QY3Y47 ID Q93Y47 PRELIMINARY; PRT; 259 AA.
AC Q93Y47;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 3-beta-hydroxysteroiddehydrogenase (EC 1.1.1.51).
GN 3HSD.
OS Digitalis lanata (Foxglove).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Veronicaceae; Digitalis.
```

```
OX NCBI_TaxID=49450;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DRESDNER; TISSUE=LEAF;
RX MEDLINE=200119909; PubMed=10550629;
RA Finsterbusch A., Lindemann P., Grimm R., Eckerskorn C., Luckner M.;
RT "Delta5-3beta-Hydroxysteroid dehydrogenase from Digitalis lanata
RL Ehrh.- a multifunctional enzyme in steroid metabolism?";
RN Planta 209:478-486(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DRESDNER; TISSUE=LEAF;
RA Thoringen C., Teuber M., Pangert A., Finsterbusch A., Lindemann P.;
RT "Spatial Organisation of Pregnanolone biosynthesis in Digitalis lanata
RL Ehrh.";
RN Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DRESDNER; TISSUE=LEAF;
RA Finsterbusch A.;
RT "Reinigung und Charakterisierung der Delta5-3beta-Hydroxysteroid
RT Dehydrogenase (Delta5-3beta-HSD) aus Digitalis lanata
RT EHRH Dissertation.";
RL Thesis (1999), Department of Institut fuer Pharmazeutische Biologie,
RL Martin Luther Universitaet Halle/Wittenberg, Halle, Germany.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=DRESDNER; TISSUE=LEAF;
RA Lindemann P., Finsterbusch A., Pangert A., Luckner M.;
RT "Partial cloning of a delta5-3-beta-hydroxysteroid dehydrogenase from
RT Digitalis lanata.";
RL (In) Okamoto M., Ishimura Y., Nawata H. (eds.);
RL Molecular Steroidogenesis, pp.333-334, Universal Academy Press, Tokyo,
RL JAPAN (2000).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL; AJ345026; CAC93667.1; -.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PROSITE; P500061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
FT CHAIN 2 259 3-BETA-HYDROXYSTEROIDDEHYDROGENASE.
SQ SEQUENCE 259 AA; 26988 MW; A782FAD6AED65B11 CRC64;

Query Match 42.6%; Score 593.5; DB 10; Length 259;
Best Local Similarity 50.2%; Pred. No. 4.1e-37;
Matches 127; Conservative 36; Mismatches 87; Indels 3; Gaps 3;

QY 10 RLEGKVALITGGASGIGETTAKLFSQHGAKVAIADVQDELGHSHVVEAIGTSNS-TYIHC 68
Db :|||||:|||||: |||: |||: |||: |||: |||: |||: |||
6 RLEGKVALITGAASGIGETARLFVEHGASVVADVQDELGRQVAVASNSDDKISYVHCD 65
QY 69 VTNEDGVKNVDMVTSTYKGLDIFMFSNAGISDPNRPRIIDNEKADFERVLSVNVTVGVFLC 128
Db :||| : | : | : |||: |||: |||: |||: ||| : ||| : ||| : |||
66 VRDEKQVAATVRVAVEKYGRKLDIMLSNAGVFGALMTNVLDLMDVDFENVLATNRGVANT 125
QY 129 MKHAARVMIPAR-SGNIISTASLSSTMGSSSHAYCGSKHVALALTRNLAVELGQFGIRV 187
Db :|||| : : : ||| : |||: ||| : ||| : ||| : ||| : ||| : |||
126 IKHAARVMVEGKVGKSGIICTSASVSASILGGMGPAYTASKHAVLGLVKGACAEGLVHGIRV 185
QY 188 NCLSPFGLPTALGKFKSGIKNEEFENFVNFAGNLKPKFNFVEDVANAALYLASDBAKYV 247
Db :||| : | : | : | : | : | : | : | : | : | : | : | : | : |
186 NSVAPYGVATPMPCSAYGM-TPSQMEANNSRANLKGVLKAKHVAEAAFLASDESAYV 244
QY 248 SGHNLFDGGSFV 260
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
245 SQNTAVDGGFTV 257

Search completed: February 14, 2003, 09:21:18
Job time : 35 secs
```

[REDACTED]

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:49 ; Search time 29 Seconds
(without alignments)
390.450 Million cell updates/sec

Title: US-09-673-918A-2
Perfect score: 1393
Sequence: 1 MOLRTAFARRLEKGVALLTG.....IDGFSVCNSVIKVOYPD 273

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	588.5	42.2	336	1 TS2_MAIZE	P50160 zea mays (m
2	406	29.1	250	1 LINC_PSEPA	P50197 pseudomonas
3	396	28.4	255	1 YWFD_BACSU	P39640 bacillus su
4	381.5	27.4	256	1 Y019_THEMEA	Q56318 thermotoga
5	376	27.0	250	1 LINC_PSEPA	P50198 pseudomonas
6	366	26.3	260	1 YK02_MYCTU	Q10855 mycobacteri
7	363	26.1	246	1 LINC_THEMEA	Q9x248 thermotoga
8	357	25.6	248	1 FABG_CHLPM	Q9z8p2 chlamydia p
9	346	24.8	251	1 Y325_THEMEA	Q9wy90 thermotoga
10	345.5	24.8	261	1 YGCW_ECOLI	P76633 escherichia
11	341.5	24.5	261	1 DHG_BACSU	P12310 bacillus su
12	334	24.0	253	1 Y4WP_RHISN	P55575 rhizobium s
13	332.5	23.9	248	1 FABG_CHLMD	Q9pkf7 chlamydia m
14	332	23.8	263	1 UCFA_SALTY	P37441 salmonella
15	328.5	23.6	256	1 DHSO_RHOSH	Q59787 rhodobacter
16	326.5	23.4	247	1 FABG_CHLTR	P38004 chlamydia t
17	320.5	23.0	244	1 FABG_VIBCH	Q9kdh7 vibrio chol
18	320.5	23.0	262	1 YXBG_BACSU	P46331 bacillus su
19	320	23.0	255	1 2BHD_STREX	P19392 streptomyce
20	316.5	22.7	244	1 FABG_ECOLI	P25716 escherichia
21	315.5	22.6	261	1 DHG_BACME	P40288 bacillus me
22	315	22.6	270	1 DHMA_FLAS1	P22441 flavobacter
23	314.5	22.6	261	1 DHG1_BACME	P39482 bacillus me
24	313.5	22.5	241	1 FABG_RICPR	P50941 rickettsia
25	310	22.3	263	1 UCFA_ECO57	Q8xbj4 escherichia
26	310	22.3	263	1 UCFA_ECOLI	P37440 escherichia
27	309.5	22.2	261	1 DHGA_BACME	P10528 bacillus me
28	308.5	22.1	244	1 FABG_SALTY	O85141 salmonella
29	308.5	22.1	261	1 DHG2_BACME	P39483 bacillus me
30	308	22.1	261	1 DHBA_BACSU	P39071 bacillus su
31	307.5	22.1	244	1 FABG_VIBHA	P55336 vibrio harv
32	306.5	22.0	261	1 DHG3_BACME	P39484 bacillus me
33	306.5	22.0	261	1 DHG4_BACME	P39485 bacillus me

ALIGNMENTS

```
RESULT 1
TS2_MAIZE
ID TS2_MAIZE STANDARD; PRT; 336 AA.
AC P50160;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Sex determination protein tasselseed 2.
DS
GN TS2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. W22;
RX MEDLINE=93364991; PubMed=8358795;
RA Delong A., Calderon-Urrea A., Dellaporta S.L.;
RT "Sex determination gene TASSELSEED2 of maize encodes a short-chain
RT alcohol dehydrogenase required for stage-specific floral organ
RT abortion.";
RL Cell 74:757-768(1993).
CC -!- FUNCTION: REQUIRED FOR STAGE-SPECIFIC FLORAL ORGAN ABORTION.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L20621; AAC37345.1; --
CC HSSP: P19992; LHDC.
CC -----
CC MaizeDB; 56963; --
CC InterPro: IPR002198; ADH_short.
CC Pfam: PF00106; adh_short; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC Oxidoreductase; Developmental protein.
KW NP_BIND 59 83 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 207 207 BY SIMILARITY.
SQ SEQUENCE 336 AA; 35204 MW; 4E273D6152B0BB99 CRC64;
Query Match 42.2%; Score 588.5; DB 1; Length 336;
Best Local Similarity 42.2%; Pred. No. 7.5e-40;
Matches 122; Conservative 50; Mismatches 88; Indels 29; Gaps 4;
QY 5 TAFARLEKGVALLTGASIGETTAKLFSQHKAKVAIADVDQLGHSVVVEAIGTSNSTY 64
Db 47 TPKPKLDGKVAIVTGGARGIGAEIVLRFKAGHGVIAIDIDDAAGEALASALGPQVS-F 105
QY 65 IHCDVTNEDGVKNAVDNTVSTY-GKLDIMFSNAGI---SDPNRPRIIDNEKADFERVLVS 120
```

P51831 bacillus su
Q51576 pseudomonas
P25529 escherichia
P42317 bacillus su
Q00791 emericella
P07914 eubacterium
P07999 bacillus me
P73574 synecocyst
P19337 eubacterium
P50199 gluconobact
P50161 aspergillus
P50842 bacillus su

34 305.5 21.9 246 1 FABG_BACSU
35 305 21.9 255 1 YV06_PSEAE
36 303 21.8 255 1 HDHA_ECOLI
37 302 21.7 257 1 YXJF_BACSU
38 301 21.6 264 1 STCU_EMENI
39 300 21.5 249 1 BA71_EUBSP
40 296 21.2 262 1 DHGB_BACME
41 295.5 21.2 247 1 FAG1_SYNY3
42 295 21.2 249 1 BA72_EUBSP
43 293 21.0 256 1 GNO_GLUOX
44 293 21.0 262 1 VER1_ASPPA
45 291 20.9 254 1 KDUD_BACSU


```
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73124; CA51638.1; -.
DR EMBL; 299123; CAB15799.1; -.
DR PIR; S39737; S39737.
DR HSSP; P50162; IAE1.
DR Subtilist; BG10628; ywfd.
DR InterPro: IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
FT NP_BIND 11 33 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 255 AA; 27324 MW; 20AA2259BFB88C9B CRC64;

Query Match 28.4%; Score 396; DB 1; Length 255;
Best Local Similarity 36.6%; Pred. No. 1.2e-24;
Matches 94; Conservative 48; Mismatches 99; Indels 16; Gaps 4;

QY 11 LGGKVALITGGASGIGETTAKLFSQHGAKVATADVODELGHSVVEAIGTSNITYIHCQVT 70
DB 5 LDKTVLITGGASGIGYAAQVAFLGQQANVVVADIDEAQGEAMVRKKNDRLHFVQTDT 64
QY 71 NEDGVKNAVDTVSTYCKLDMFSNAGIS--DPNRPRIIDNEKADFERVLSVNVTVGVFLC 128
DB 65 DEACQAHVESAVHTFGGLDVLINNGAIEIVAP-----IHEMLSDMNKVLQVNLTMGMFLM 120

QY 129 MKHAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAVLALTRNLAVELGQFGIRVN 188
DB 121 SKHALKHLMAAGCKNIINTSCVSGVLVWPDPAYNASKGVQLQIKSMADVAKHQIRVN 180

QY 189 CLSPFGLPTALGKFKSGIKNEEFENV-----INFAGNLKGPKNVEDVANAALYLASD 242
DB 181 CVCPGIDTPLNEKSFLENNEGLEEKKEKAVNPVLLRLGKP----EBIANVMLFLASD 236

QY 243 EAKYVSGHNLFDGGFSV 259
DB 237 LSSYWTGSAITADGGYT 253

RESULT 4
Y019.THEMA
ID Y019.THEMA STANDARD; PRT; 256 AA.
AC Q56318;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative oxidoreductase TM0019 (EC 1.-.-.-).
GN TM0019.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=96125254; PubMed=8550425;
RA Kletzlin A., Adams M.;
RT "Molecular and phylogenetic characterization of pyruvate and 2-
RT ketoisovalerate ferredoxin oxidoreductases from Pyrococcus furiosus
RT and pyruvate ferredoxin oxidoreductase from Thermotoga maritima.";
RL J. Bacteriol. 178:248-257(1996).
```

```

RN Nature 390:249-256(1997).
RC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X85171; CA559459.1; -.
DR EMBL; AE001690; AAD35113.1; -.
DR HSSP; O70351; IE6W.
DR TIGR; TM0019; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; NADP; Complete proteome.
FT NP_BIND 9 33 NADP (BY SIMILARITY).
FT ACT_SITE 153 153 BY SIMILARITY.
FT CONFLICT 130 136 RGGVII -> TRWRSDH (IN REF. 1).
SQ SEQUENCE 256 AA; 28078 MW; D68160BID7980C6B CRC64;

Query Match 27.4%; Score 381.5; DB 1; Length 256;
Best Local Similarity 35.7%; Pred. No. 1.8e-23;
Matches 91; Conservative 54; Mismatches 99; Indels 11; Gaps 5;

QY 11 LGGKVALITGGASGIGETTAKLFSQHGAKVATADVODELGHSVVEAIGTS--NSTYIHC 68
DB 2 LGGKVAVTGGGGIGAGAAQLFAENGKVKVIAEIDEEAGVEREMLRGLDVTVEKTD 61
QY 69 VTNEDGVKNAVDTVSTYCKLDMFSNAGISDPNRPRIIDNEKADFERVLSVNVTVGVFLC 128
DB 62 VADENSVRKNVRKTVIEYGGVDVLVNNAAVMVK--SIFERPLEEWEVRVNLMTGPYIC 119

QY 129 MKHAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAVLALTRNLAVELGQFGIRVN 188
DB 120 SRYCAEEMIKRGGVYIINIASTRAQSEPDTPYSASKGLVALTHSLAVSLRYHVRV 179

QY 189 CLSPFGLPTALGKFKSGIKNEE--EFENYINPAGNLKGPKNVEDVANAALYLASDE-AK 245
DB 180 SISPGWIETSEWKKSLRKKPDLRPIDHQHPAGRVGNPL----DIAHLCLVFLADDEKAG 235

QY 246 YVSGHNLFDGGFSV 260
DB 236 FITGTNFIVDGGMTV 250

RESULT 5
LINX.PSEPA
ID LINX.PSEPA STANDARD; PRT; 250 AA.
AC P50198;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase (EC 1.1.-.-)
GN LINX.
```


QY	127	LCWKHAARVMIPARSGNIISTASLSTMTGGSSSHAYCGSKHVAVLATRLNLADELGQFGIR	180
Db	119	NVTQMVVPYMKORNGSIVNVSSVGGIYNPQGTNAASKAGVIGTKTWAKELAGNRIR	178
QY	187	VNCLSPGLPTALGKFKSGIKNEEENENFINFAGNLKGFKNVEDVANAALYLASDEAKY	246
Db	179	VNAVAPGFIETPMTEKLPEKARETALSRIIP--LGRFGKP-----EEVAQVILFLASDESSY	232
QY	247	VSGHNLFIDGGFSV	260
Db	233	VTQGVIGIDGGLVI	246
RESULT 8			
FABG_CHLPN			
ID	FABG_CHLPN	STANDARD;	PRT; 248 AA.
CD	Q9Z8P2; O9J0D1;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl		
DE	acyl carrier protein reductase).		
GN	FABG OR CPN0296 OR CP0462.		
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
OX	NCBI_TaxID=83538;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CWL029;		
RX	MEDLINE=99206606; PubMed=10192388;		
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,		
RA	Olinger L., Grinwood J., Davis R.W., Stephens R.S.;		
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";		
RT	Nat. Genet. 21:385-389(1999).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=AR39;		
RX	MEDLINE=20150255; PubMed=10684935;		
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,		
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,		
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,		
RA	Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.		
RA	Eisen J., Fraser C.M.;		
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia		
RT	pneumoniae AR39.";		
RT	Nucleic Acids Res. 28:1397-1406(2000).		
RP	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=J138;		
RX	MEDLINE=20330349; PubMed=10871362;		
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,		
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;		
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138		
RT	from Japan and CWL029 from USA".;		
RT	Nucleic Acids Res. 28:2311-2314(2000).		
CC	-1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +		
CC	NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.		
CC	-1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS		
CC	PATHWAY.		
CC	-1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES		
CC	(SDR) FAMILY.		
This SWISS-PROT entry is copyright. It is produced through a collabora			
between the Swiss Institute of Bioinformatics and the EMBL outstat			
the European Bioinformatics Institute. There are no restrictions on			
use by non-profit institutions as long as its content is in no			
modified and this statement is not removed. Usage by and for comm			
entities requires a license agreement (See http://www.isb-sib.ch/anno			
or send an email to license@isb-sib.ch).			
CC	EMBL; AE001614; AAD18445.1; -		
DR	EMBL; AE002207; AAF38299.1; -		


```

RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RT DNA Res. 4:91-113(1997).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U29579; AAA69285.1; ALT_INIT.
DR EMBL: AE000361; AAC75816.1; ALT_INIT.
DR EMBL: D90893; BAA16569.1; ALT_INIT.
DR HSSP: P25529; 1AHH.
DR Ecogene: EGI3130; ygcW.
DR InterPro: IPR002198; ADH_short..
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
FT NP_BIND 22 46 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 166 166 BY SIMILARITY.
SQ SEQUENCE 261 AA; 28042 MW; 2D6B2382768A4F7E CRC64;

Query Match 24.8%; Score 345.5; DB 1; Length 261;
Best Local Similarity 32.3%; Pred. No. 1.4e-20;
Matches 82; Conservative 58; Mismatches 101; Indels 13; Gaps 6;

QY 11 LRGKVALITGGASGIGETAKLFQSHGAKVAIADVQDELGHS--VVEAIGTSNSTVIHCD 68
DQ 16 LRGKTAIVTGGNSGLGQAFAMALAKAGANIFTPSPKDNGETKEMIEKGV-EVDFMQVG 74
QY 69 VTNEDGVKNADNTVSTYKGLDIFSNAGISDPNRIIDNEKADFERVLSVNVTVGVFLC 128
DQ 75 ITAEGAPQRIIAACCBFRCTGVTDILVNNAGICKLN--KVLDGFRADWDPMIDVNLTAAPFL 132
QY 129 MKHAARVMTPARSGNITISASTSMTGGSSHAYCGSKHAVLALTENLAVELGQFGIRVN 188
DQ 133 STEAKIMIPQSGKIINCSLFSYLGQWSPAYSATKHALAGFTRKAYCDELGQYNIQVN 192
QY 189 CLSP--FGLPTALGKFKSGIKNEEFENVFNAGNLKPKFNVEDVANAALYLASDEARY 246
DQ 193 GIAPGYATDITLATSNFETNORVLDHI---PANRWG---DTQDLMGRAVFLASPASNY 246
QY 247 VSGHNLFDIGGFSV 260
DQ 247 VNGHLLVVDGGYLV 260

RESULT 11
DHG_BACSU STANDARD; PRT; 261 AA.
AC P12310; P94430;
DT 01-OCT-1989 (Rel. 12, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose 1-dehydrogenase (EC 1.1.1.47).
GN GDH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86168021; PubMed=3082854;
RT Lampel K.A., Uratani B., Chaudhry G.R., Ramaley R.F., Rudikoff S.;
RT "Characterization of the developmentally regulated Bacillus subtilis
RT glucose dehydrogenase gene."
RL J. Bacteriol. 166:238-243(1986).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis chromosome:
RT determination of the sequence of a 146 kb segment and identification
RT of 113 genes."
RL Microbiology 142:3047-3056(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chol S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Katamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- CATALYTIC ACTIVITY: Beta-D-glucose + NAD(P)(+) = D-glucono-1,5-
CC lactone + NAD(P)H.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- INDUCTION: IT IS INDUCED AT STAGE III OF THE SPORULATION.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M12276; AAA22463.1; -.
DR EMBL: D50453; BAA09024.1; -.
DR EMBL: Z99106; CAB12201.1; -.
DR PIR: S36090; S36090.
DR HSSP: P50162; 1AEL.
DR Subtilist: BG10545; qdh.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NADP; Sporulation; Complete proteome.
FT NP_BIND 11 35 NADP (BY SIMILARITY).
FT ACT_SITE 158 158 BY SIMILARITY.
FT CONFLICT 148 150 EVI -> AF (IN REF. 1).
SQ SEQUENCE 261 AA; 28090 MW; 5894C17DB8F14965 CRC64;

```

Query Match	24.5%;	Score	341.5;	DB 1;	Length	261;			
Best Local Similarity	24.4%;	Pred.	No. 2.8e-20;						
Matches	93;	Conservative	50;	Mismatches	98;	Indels	29;	Gaps	10
QY	11	LEGGKVALITGASGIGETTKAKLFSOHGAKVAI--ADYQD--ELGHSVVEAIGTSNSTYI	65						
DB	5	LKGKVAITGAASGLGKAMAIREFGEOAKVINYYSNKQDPNEYKEEVIKAGG--EAVVV	62						
QY	66	HCDVTNEDGVKNADVNTVSYGKLDLIMFSNAGISDPNPRIIDNEK--ADPERVLSVNWIT	123						
DB	63	QGDVTKEDYKNIIVQTAIKBEGFLDIMIINAGLENP---VPSEHMLPKDWKVGIGTNL	118						
QY	124	GVFLCMKHAARVMIPAR--SGNIITASTSSTTGGCGSSHAYCGSKSHAVLATRNIAVLGQ	182						
DB	119	CAGLGREAKYIFVENDIKGNINMSSEVHEIPLWPLFVHYAASGGGKTKMTETIALEYAP	178						
QY	183	FGIRVNCLSDFGLPTAL--GKKFSGKNEEFPNTNFAGNLKPKPFNVEDVANAALYLS	241						
DB	179	KGIRVNIIGPGLNTPLNAEKFDPKQADVESMPLM--GYIGEP-----ETIAVAAWLAS	233						
QY	242	DEAKYVSGHNLFDGGGFSVCNSVKKVQYFP	271						
DB	234	KEASVYTGITLFDAGGMT-----QYF	254						

```

RESULT 12
Y4MP_RHISN
ID Y4MP_RHISN STANDARD; PRT; 253 AA.
AC P55575;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative short-chain type dehydrogenase/reductase Y4MP (EC 1.-.-.-).
GN Y4MP.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
NCBI_TaxId=394;
[1]
SEQUENCE FROM N.A.
MEDLINE=97303956; PubMed=9163424;
Freiberg C.A., Fellay R., Balroch A., Broughton W.J., Rosenthal A.,
Perist X.;
Molecular basis of symbiosis between Rhizobium and legumes.";
Nucleic 367:1394-401(1997).
-1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY. HIGH, TO BACTERIAL FABG.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; AE000085; AAB91779.1; -
HSSP; E29132; 1DFI.
InterPro; IPR002198; ADH_short.
Pfam; PF00106; adh_short; 1.
PROSITE; PS00061; ADH_SHORT; 1.
Hypothetical protein; Oxidoreductase; Plasmid.
NP_BIND 10 35 NAD OR NADP (BY SIMILARITY).
ACT_SITE 159 159 BY SIMILARITY.
FT ACT_SITE 159 159
SEQUENCE 253 AA; 25994 MW; B8208A0B0F542F CRC64;
Query Match 24.0%; Score 334; DB 1; Length 253;
Best Local Similarity 34.2%; Pred. No. 1.1e-19;
Matches 90; Conservative 55; Mismatches 92; Indels 26; Gaps 10;
OY 11 LEGKVALITGGAS--GIGETAKLFSGHGAKVADVQ--DELGHSVWEAIGTSTVI-- 65

```

Db	3	LGKTAVLSGAASKRGTRATELFAHGARYALLDINADEKAAAGDLPVPEHAGHGL	62
Qy	66	HCDDVNECDVKNVAVNTVSTYTGKLDITMFSNAGISODPNRPITDNEKADFERYLSVNTGV	125
Db	63	RCDDVADRASCT'SASDEVLSAFGVANILINNAGITQP-VKTLDISADQWORIYAVNKTGV	120
Qy	126	FLCMKHAARVMIPAR---SGNITISTASLSSTMGGG---SSHAYCGSKHAVILATRLNAYE	179
Db	121	L-----NLGVEITPNMRONGGGSITACMSSVSAQGGGIEGPHYSNAKAGVILGLAKAVARE	176
Qy	180	LQGFGRIVNCILSPFGLTAL-GRKFGSGIRNEEFENV-INFAGNLKGPKENVEDVANAAL	237
Db	177	FGPDSIRVNCVTPGLIOTDTGTGKLSAEMRADIVKGIPLSLRGDAR-----DVANIYL	229
Qy	238	YLASDEAKYVSGHNLFIDGGFSV	260
Db	230	FLASDLSAYVTGAVIDVNGMLI	252

RESULT 13	FABG_CHLMU	STANDARD;	PRT;	248 AA.
AD	Q9PWF7; 2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-			
DE	acyl-CoA thiolase protein reductase)			
GN	FABG OR TC0506			
OS	Chlamydia muridarum			
OS	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
NCBI	NCBI_TaxID=83560;			
XX	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=MopN / N19g;			
RX	MEDLINE=20150255; PubMed=10684935;			
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,			
RA	White O., Hackey E.K., Peterson J., Utkerback T., Berry K., Bass S.,			
RA	Linhner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,			
RA	Glinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,			
RA	Eisen J., Fraser C.M.;			
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia			
RT	pneumoniae AR39.";			
RL	Nucleic Acids Res. 28:1397-1406(2000).			
CC	-!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +			
CC	NADPH(+) -> 3-oxoacyl-[acyl-carrier protein] + NADPH.			
CC	-!- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS			
CC	PATHWAY.			
CC	-!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES			
CC	(SDR) FAMILY.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collabora-			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no			
CC	modified and this statement is not removed. Usage by and for commer-			
CC	entities requires a license agreement (See http://www.isb-sib.ch/annou-			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; AE002319; RAF39350.1; -			
CC	HSP; P50163; 2AE1.			
DR	TIGR; TC0508; -			
DR	InterPro; IPR02198; ADH_short.			
DR	Pfam; PF00106; adh_short; 1.			
DR	PRINTS; PRO0080; SDRFAMILY.			
DR	PROSITE; PS00061; ADH_SHORT; 1.			
DR	Fatty acid biosynthesis; Oxidoreductase; NADP; Complete proteome.			
DR	NP_BIND 12 36			
DR	ACT_SITE 157 157			
DR	SEQUENCE 248 AA; 25977 MW; 1F5C8968CB05F58 CRC64;			
DR	TSQ			

Query Match 23.9% Score 332.5 DB 1; Length 248;

```

Best Local Similarity 32.9%; Pred. No. 1.4e-19;
Matches 85; Conservative 52; Mismatches 98; Indels 23; Gaps 7;
QY 11 LECKVALITGGAGSIGETAKLFQSHGCAKVAIADVOELGSHVVEAIGT---SNSTYIHC 67
Db 5 LVNKAIVTGGSGIGFGIAKFAEHGANVQIWGINEEAGKSAQDLSDKTSKVSFALV 64
QY 68 DVTNEDGVKNADVNTVSTYCKLDMFSNAGIS-DPNRPRIIDNEKADFERVLSVNTGVF 126
Db 65 DYSKNDMSVAQVQKFLAEYGTIDVVNNAGITRDSLMLRMSEEE---WSVIDTNLGSIY 121
QY 127 LCMKHARVWPARSNIISTASLSMTGSGSHYCGSKHVAIATRLNLAVELGQFGIR 186
Db 122 NVCSAVIRPMIKARSAIYNISSIVGLRSGPGQTNTYAAAKAGIIGFSKALSKEVGSKNIR 181
QY 187 VNCLSPFGLPTALGKFFS-GIKNEEEFENFVINFAGNLKG-PRFNV---EDVANAALYLAS 241
Db 182 VNCIARGFDITDMTKLSNLKNW-----LKGVLGRVGTPEIAMAALFLAS 230
QY 242 DEAKYVSGHNLFTDGGFS 259
Db 231 NQSSYITGVLSVDGMA 248

RESULT 14
UCPA_SALTY STANDARD; PRT; 263 AA.
AC P37441;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Oxidoreductase ucpa (EC 1.-.-.-).
GN UCPA OR STM2445 OR STY2682.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
ON NCBI_TaxID=602, 601;
RX SPECIES FROM N.A.
RC SPECIES-S typhimurium; STRAIN=LT2 / SGSCI412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 204-263 FROM N.A.
RC SPECIES-S typhimurium;
RX MEDLINE=91358382; PubMed=1909324;
RA Hryniewicz M.M., Kredich N.M.;
RT "The cysp promoter of Salmonella typhimurium: characterization of two
RT binding sites for CysB protein, studies of in vivo transcription
RT initiation, and demonstration of the anti-inducer effects of
RT thiosulfate."
RL J. Bacteriol. 173:5876-5886(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-S typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher K., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davis R.M., Dowd L., White N., Farrar J.,
RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella

```

```

RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE008810; AAL21339.1; -
DR EMBL; AL627274; CAD07676.1; -
DR PIR; A38121; A38121
DR StyGene; SG10424; ucpa
DR InterPro; IPR002198; ADH_short.
DR PROSITE; PS00061; ADH_SHORT; FALSE_NEG.
KW Oxidoreductase; Complete proteome.
FT CONFLICT 218 L -> M (IN REF. 2).
FT CONFLICT 258 S -> T (IN REF. 2).
FT CONFLICT 263 V -> I (IN REF. 2).
SQ SEQUENCE 263 AA; 27870 MW; 81B26CAD9B4D534E CRC64;

Query Match 23.8%; Score 332; DB 1; Length 263;
Best Local Similarity 33.1%; Pred. No. 1.6e-19;
Matches 87; Conservative 49; Mismatches 113; Indels 14; Gaps 5;
QY 10 RLEGKVALITGGAGSIGETAKLFQSHGAKVAIADVOELGSHVVEAIGTSN-STYIHC 68
Db 3 KLTGTALITGASQIGEGIAVFARHGANLILDISDEIKELADELGGHRCRTAVKAD 62
QY 69 VTNEDGVKNADVNTVSTYCKLDMFSNAGISDPNRPRIIDNEKADFERVLSVNTGVFLC 128
Db 63 VRDFASVQAAVARAKETEGRIDILVNNAGVRLGNFLDMSEEDRPH--IDINIKGMVW 120
QY 129 MKHAARVMIPARSGNIISTASLSSTM-GGSSHAYCGSKHVAIATRLNLAVELGQFGIRV 187
Db 121 TRAVLPMLKRDGRIVMMSSVTGDMVADPGETAYALSAAIIVGLTKSLAVEYAQSGIRV 180
QY 188 NCLSPFGLPTALGKFFSGIKNEEEFENVINFAG-----NLKGPKNFVEDVANAALYLAS 241
Db 181 NAICPGYVTRPNAESTARQSNPDPSVLTEMAKAIPRLRLADPL----EVGLAAFLAS 236
QY 242 DEAKYVSGHNLFTDGGFSVCNV 264
Db 237 DESSYLTTGQNVIDGGSTLPESV 259

RESULT 15
DHSD_RHOSH STANDARD; PRT; 256 AA.
AC Q59787;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase)
DE (Polyol dehydrogenase).
GN POLS OR SMOS.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=96032011; PubMed=7551049;
RA Schauder S., Schneider K.-H., Giffhorn F.;
RT "Polyol metabolism of Rhodobacter sphaeroides: biochemical
RL characterization of a short-chain sorbitol dehydrogenase."
RL Microbiology 141:1857-1863(1995).
CC -!- FUNCTION: CATALYZES THE OXIDATION OF D-GLUCITOL (SORBITOL) TO D-

```

```

CC PROCTOSE, GALACTITOL TO D-TAGATOSE AND OF L-IDITOL.
CC -!- CATALYTIC ACTIVITY: L-Iditol + NAD(+) = L-sorbose + NADH.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF018073; AAC45770.1; -.
CC DR HSSP: PI9992; 1HDC.
CC DR InterPro: IPR002198; ADH_short.
CC DR Pfam: PF00106; adh_short; 1.
CC DR PRINTS: PR00080; SDRFAMILY.
CC DR PROSITE: PS00061; ADH_SHORT; 1.
CC DR Oxidoreductase; NAD.
CC KW ACETALDEHYDE 5 34 NAD (BY SIMILARITY).
CC FT ACT_SITE 152 152 BY SIMILARITY
CC FT SEQUENCE 256 AA; 27014 MW; 33CD7BC17598B02A CRC64;
CC -----
Query Match 23.68; Score 328.5; DB 1; Length 256;
Best Local Similarity 32.08; Pred. No. 3e-19;
Matches 83; Conservative 52; Mismatches 105; Indels 19; Gaps 7;

QY 10 RLEGKALITGASIGETTKAKLFSQHGAKVAIVADQDELGHSVVEATGTSNSTYHCDV 69
DB 1 :||| ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
2 RLDGKTALLTGARGIGRAFAEAVREGARVAIADINLEAARATAEATGPA-ACAALDV 60
DB 1 :||| ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 70 TNEDGKYNADVNTYSTYCKLDMFSNAGISDPNRPRIIDNEKADFERVLSVNTGVFLCM 129
DB 1 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
61 TDOASIDRCVAFELDRMGSIDILVNNALFD--LAPIVETRESTDRFAINVSTLFMM 118
DB 1 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 130 KHAARVMYIP-ARSGNIISTASLSTMTGGGSHAYCSCSHAVIALTRNLAVELGFGIRVN 188
DB 1 :||| ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
119 QAVARAMTAGRGKKIINMASQGRGEALVGYVCATKAIVSLTQSGAGLNLIIRHGINVN 178
DB 1 :||| ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 189 CLSPFGPLTALCKKFPSTKINP-EEFENVINFAGNLK-----GPKFNVEDVANAALY 238
DB 1 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
179 AIAP---GVVGEHWDGVDKAFADYENLPR--GEKKRVGAAVPFGRMGRAEDLTGMAIF 233
DB 1 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 239 LASDEAKTVSYSGHNLFIDGG 257
DB 1 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 234 LATPEADYIVAOTYNVDGG 252
DB 1 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:54 ; Search time 19 Seconds
(without alignments)
1381.300 Million cell updates/sec

Title: US-09-673-918A-2
Perfect score: 1393
Sequence: 1 MQLRTAFARLEKGVALLTG.....IDGGFSVCNSVIRKVFQYDPS 273
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	734	52.7	267	T11579	probable short cha
2	685	49.2	284	T02257	probable short cha
3	658.5	47.3	285	F95563	hypothetical prote
4	625.5	44.9	303	T46064	short-chain alcoho
5	597.5	42.9	283	H85039	probable alcohol d
6	588.5	42.2	336	A47542	short-chain alcoho
7	575.5	41.3	264	T02174	probable alcohol d
8	567	40.7	257	T02175	probable alcohol d
9	559	40.1	251	T06364	probable short-cha
10	549.5	39.4	258	T02176	probable alcohol d
11	499.5	35.9	234	T03734	short chain alcoho
12	490.5	35.2	272	T47354	alcohol dehydrogen
13	432	31.0	253	F83440	probable short-cha
14	402.5	28.9	258	C70885	probable dehydroge
15	398	28.6	256	A10406	probable dehydroge
16	396	28.4	255	S39737	glucose 1-dehydrog
17	386	27.7	255	G82644	2,5-dichloro-2,5-c
18	381.5	27.4	256	E72427	oxidoreductase, sh
19	370.5	26.6	260	T36846	probable dehydroge
20	366	26.3	260	H70758	probable fabG3 pro
21	363	26.1	246	H72219	3-oxoacyl-(acyl ca
22	363	26.1	254	AC1478	dehydrogenase/redu
23	363	26.1	254	A11117	dehydrogenase/redu
24	361.5	26.0	296	E87260	hypothetical prote
25	357	25.6	248	H86527	oxoacyl (carrier p
26	357	25.6	248	H72096	3-oxoacyl-(acyl-ca
27	356.5	25.6	243	F83838	oxidoreductase (sh
28	356	25.6	246	H84136	3-oxoacyl-(acyl-ca
29	353.5	25.4	286	F91082	probable oxidoredu

ALIGNMENTS

RESULT 1

T11579
probable short chain alcohol dehydrogenase CPRD12, drought-inducible - cowpea
C:Species: Vigna unguiculata (cowpea)
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 20-Jun-2000
C:Accession: T11579
R:Uchi, S.; Yamaguchi-Shinozaki, K.; Urao, T.; Shinozaki, K.
J. Plant Res. 109, 415-424, 1996
A:Title: Characterization of two cDNAs for novel drought-inducible genes in the highl
A:Reference number: Z17293
A:Accession: T11579
A:Molecule type: mRNA
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-267 <IUC>
C:Cross-references: EMBL:D88121
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 52.7%; Score 734; DB 2; Length 267;
Best-Local Similarity 56.3%; Pred. No. 8.3e-51;
Matches 147; Conservative 41; Mismatches 67; Indels 6; Gaps 2;

QY	9	RRLEKVALITGGASGIGETTAKLFQSHGKVAIAADVQDELGHSVVEAIGTSNITYHCD 68
DB	12	KRLGKVAITGGASGIGETARLFQSHGAVVVAIDQDDVGLSLCNEL--KSAIVVHCD 69
QY	69	VTNEDGVKNADVNTVSTYKLDIMFSNAGISDPNRPRIIDNEKADFERVLSVNVGTGFLC 128
DB	70	VTKEEDIERKCVDTAVSKFGKLDIMFNAGTGDEFKKSILDNTKSDFERVISLVNLPFLG 129
QY	129	MKHAARVMIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLATRNLAVELGQFGIRVN 188
DB	130	THAARVMIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLATRNLAVELGQFGIRVN 189
QY	189	CLSPFGLTALGKFKSGIKNEEFENVINFAGNLKPKFNVEDVANAALYLASDEAKYVS 248
DB	190	CVSPFAIVTPLLKNYENLDEGVRKTYMNLKGWYVPV---NDVAEALYLASDESKEVS 245
QY	249	GHNLFDGGFSVCNSVIRKVFQ 269
DB	246	SHNLVIDGGLINSNVGPFMFE 266

RESULT 2

T02257
probable short chain alcohol dehydrogenase - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000
C:Accession: T02257
R:Herbers, K.; Moenke, G.; Badur, R.; Sonnewald, U.
Plant Mol. Biol. 25, 1027-1038, 1995
A:Title: A simplified procedure for the subtractive cDNA cloning of photoassimilate-r
A:Reference number: S62698; MUID:96145513; PMID:8555446

A:Accession: T02257
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-284 <HER>
A:Cross-references: EMBL:AJ223177; NID:g2739278; PIDN:CAA11153.1; PID:g2739279
A:Experimental source: Cultivar SNN; tissue-type leaf
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 49.2%; Score 685; DB 2; Length 284;
Best Local Similarity 53.8%; Pred. No. 6.9e-47;
Matches 140; Conservative 37; Mismatches 83; Indels 0; Gaps 0;

QY 3 LRAPARLEKVALITGGASGETTAKLFSQHGAKVAIADVDQDELGHVSVEAIGTSNS 62
DB 6 LPSIAKLEKVALITGGASGETTAKLFSQHGAKVAIADVDQDELGHVSVEAIGTSNS 65
QY 63 TYIHCVDVTNEDGVKNVNTVSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSNV 122
DB 66 IFVHCNVAVESDVQNVYDIAKFLGDLIMFMSNAGIGKSSISLDYDIKTQVFDVNI 125
QY 123 TGVFLCMKHAARVMIPARSNGIISTASLSMTGGSSHAYCGSHAVLALTRNLAVELGQ 182
DB 126 VGEFFCAKHAARVMIPKKSIIFTASATVSVGVPHAYSASKGAVLGFSGKNGVELGK 185
QY 183 FGIRVNCLSFPLTALGKFGSKKEEFENVINFAGNLGKPFNFVEDVANAALYLASD 242
DB 186 YGIRVNCVPHYSTPLVNLALGIAEITAERKFAAGNLGALLDEEVAKAVLYLASD 245
QY 243 EAKYSGHNLFIDGGFSCVN 262
DB 246 DSKYSGMNLVIDGGFSTTN 265

RESULT 3
F96563
Hypothetical protein F96563 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
A:Accession: F96563
R:Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: F96563
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-285 <STO>
A:Cross-references: GB:AE005173; NID:g10645436; PIDN:AAG21552.1; GSPDB:GN00141
C:Genetics:
A:Gene: F96563
A:Map position: 1
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 47.3%; Score 658.5; DB 2; Length 285;
Best Local Similarity 50.9%; Pred. No. 8.7e-45;
Matches 139; Conservative 43; Mismatches 78; Indels 13; Gaps 5;

QY 8 ARRLKVALITGGASGETTAKLFSQHGAKVAIADVDQDELGHVSVEAI--GTSNST-- 63
DB 15 SORLLKVALITGGATGIESIVLFLPKHGAKVCIVLDQDLGGECVCKSLRSGKETAF 74
QY 64 YIHCVDVTNEDGVKNVNTVSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSNVNT 123
DB 75 FIHGDRVEDDISNAVDFAKNFGTDLINNALGCGAPCPDIRNYSLESEFMFDVNVK 134
QY 124 GVFLCMKHAARVMIPARSNGIISTASLSMTGGSSHAYCGSHAVLALTRNLAVELGQF 183

DB 135 CAFLSMKHAARVMIPKKSIVSLCSVGVGVGPHSVGSKHAVLGLTRSVAAELQOH 194
QY 184 GIRVNCLSFPLTALGKFGSKKEEFENVI---NFA---GNLKGPKFNVDEVANAA 236
DB 195 GIRVNCVSPYAVATKLA--LAHLPEERTEDAEVFPNFAANANLKGVELTVDDVANAV 252
QY 237 LYLASDEAKYVSGHNLFIDGGFSCVNSVIKVFQ 269
DB 253 LFLASDDSRYSISGDNLMIDGFTCTNHSFKVFR 285

RESULT 4
T46064
short-chain alcohol dehydrogenase-like protein - Arabidopsis thaliana
N:Alternate names: protein T18N14.60
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Mar-2000
C:Accession: T46064
R:Delzeny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23013
A:Accession: T46064
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-303
A:Cross-references: EMBL:AL132968
A:Experimental source: Cultivar Columbia; BAC clone T18N14
C:Genetics:
A:Map position: 3
A:Introns: 31/2
A:Note: T18N14.60
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 44.9%; Score 625.5; DB 2; Length 303;
Best Local Similarity 47.5%; Pred. No. 3.9e-42;
Matches 131; Conservative 52; Mismatches 78; Indels 15; Gaps 6;

QY 5 TAFARLEKVALITGGASGETTAKLFSQHGAKVAIADVDQDELGHVSVEAIGTSNST- 63
DB 26 TLYPKRLEKVALITGGAGHIGKATVLMFAHKGATVVIADVDNVAGSSLSHKTSP 85
QY 64 ---YIHCVDVTNEDGVKNVNTVSTYKGLDIFMSNAGI--SDPNRPRIIDNEKADFERV 118
DB 86 MVAFISCDVSVEADENVLVNTVARYGRDLILFNAGVLGDQKKHKSILDFDADEFDHVM 145
QY 119 SVNVTVGLCMKHAARVMIP-ARSGNIISTASLSMTGGSSHAYCGSHAVLALTRNLA 177
DB 146 RVNVYVGLGKMGKHCARAMIKRGFKGCIISTASVAGVMGGMGPHTASKHAIVGLTKNAA 205
QY 178 VELQFGIRVNCLSFPLTAL----GKFFSG--IKNE--EEFENVINFAGNLGKPFNV 229
DB 206 CELKYGIRVNCISPFQVATSLVNLNARVKTSGDVEDDVEEVEEFVRSANLAKGETLRA 265
QY 230 EDVANAALYLASDEAKYVSGHNLFIDGGFSCVNSVI 265
DB 266 NDIAEAALYLASDESRYVNGHNLVVDGVTARNVCV 301

RESULT 5
H85039
probable alcohol dehydrogenase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
C:Accession: H85039
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: H85039
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-283 <STO>

A:Cross-references: GB:NC_001268; NID:g7270184; PIDN:CAB77799.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g03140
A:Map position: 4
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 42.9%; Score 597.5; DB 2; Length 283;
Best Local Similarity 47.0%; Pred. No. 5.9e-40;
Matches 126; Conservative 45; Mismatches 86; Indels 11; Gaps 4;

QY 9 RRLGKVALITGGASIGETTAKLFSQHGAIAADVQDELGHSVVEATGTSNTHCD 68
DB 16 KRLGKVALITGGASIGETAKGFIHQHAKVIAIDQIQIGRETQDELGPS-CAYFPD 74
QY 69 VTNEDGVKNADVNTSTYGLKIDIMFNSAGISDPNRPRIIDNEKADFERVLSVNVTVGFLC 128
DB 75 VTRESIDANADVAVSLHTKLDIMYNAGIPCKTPPSIVDLDLNVKDKYNTNVRGMAG 134
QY 129 MKHAARVMPARGSNIIISTASLSTMTGGSSSHAYCGSKHVALTLRNLAVALGQFGIRVN 188
DB 135 IKHAARVMPRNSGSIITAGSVTGMGLAQHTYSVSKSAVIGIVRSTASELCKHRIRVN 194
QY 189 CLSPFGLPTA----LGKFSGIGKNEEFENVINFAGNLKGPKNFVEDVANAAVYLASDE 243
DB 195 CISPFALITFSVDEMRQIYPGV-DSRLIQIVQSTGVNLGVECEPTDVANAAVYLASDD 253
QY 244 AKYVGNHNLFDGGFSCVNSVIRKVFQYP 271
DB 254 SKYVGNHNLVVDGGF---TTVKTLDFP 277

RESULT 6
A47542
short-chain alcohol dehydrogenase (EC 1.1.1.1) - maize
C:Species: Zea mays (maize)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 31-Mar-2000
C:Accession: A47542
R:DeLong, A.; Calderon-Urrea, A.; Dellaporta, S.L.
Cell 74, 757-768, 1993
A:Title: Sex determination gene tasselseed2 of maize encodes a short-chain alcohol dehydrogenase
A:Reference number: A47542; MUID:93364991; PMID:8358795
A:Accession: A47542
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-336
A:Cross-references: GB:I20621; NID:g393183; PIDN:AAC37345.1; PID:g393184
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; oxidoreductase
F:56-238/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 42.2%; Score 588.5; DB 2; Length 336;
Best Local Similarity 42.2%; Pred. No. 3.8e-39;
Matches 122; Conservative 50; Mismatches 88; Indels 29; Gaps 4;

QY 5 TAFARLEKVALITGGASIGETTAKLFSQHGAIAADVQDELGHSVVEATGTSNSTY 64
DB 47 TPMKRLDKVAITVGARGIGEAIVRLFAKHGARVVIADIDDAAGEALASALGPQVS-F 105
QY 65 IHCDVTNEDGVKNADVNTSTY-GKLDIMFNSAGI--SDPNRPRIIDNEKADFERVLSV 120
DB 106 VRCDVSVEDDVRRAVDWALSRHGRLDVCNNAGVILGRQTRAAARSILSFDAAEFDRLRV 165
QY 121 NVTVGFLCMKHAARVMPARGSNIIISTASLSTMTGGSSSHAYCGSKHVALTLRNLAVAL 180
DB 166 NALGAALGMKHAARMAPRAGSIVSVASVAAVLGLGPHAYTASHKHAIVTLTKNAACEL 225
QY 181 GQFGIRVNCLSPPGLPTA-----LGKFSGIGKNEEFENVNI 216
DB 226 RAHGVRVNCVSPFGVATPMLINAWRGCHDDATADARDLDDLDVTPSPDQVEKMEEV 285
QY 217 NFAGNLKGPKNFVEDVANAAVYLASDEAKYVGNHNLFDGGFSCVNSVI 265
DB 286 RGLATLKGPTRLPRDIAEAVLFUASDEARYISGHNLVVDGGVTTSRNLI 334

RESULT 7
T02174

probable alcohol dehydrogenase [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F14M4.3

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001

C:Accession: T02174; E84911

R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; submitted to the EMBL Data Library, September 1998

A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.

A:Reference number: Z14609

A:Accession: T02174

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-264 <ROU>

A:Cross-references: EMBL:AC004411; NID:g3522932; PID:g3522952

R:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: E84911

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-264 <STO>

A:Cross-references: GB:AE002093; NID:g3522952; PIDN:AAC34234.1; GSPDB:GN00139

C:Genetics:

A:Gene: F14M4.3; At2g47140

A:Map position: 2

A:Introns: 163/2

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 41.3%; Score 575.5; DB 2; Length 264;

Best Local Similarity 46.9%; Pred. No. 3e-38;

Matches 120; Conservative 49; Mismatches 82; Indels 5; Gaps 4;

QY 7 FARRLEKVALITGGASIGETTAKLFSQHGAIAADVQDELGHSVVEATGTSNSTYIH 66
DB 9 FFSRLDGKIVITGGASIGAGSVRLTEHGARVVIQVDELGQNVAVSIGEDKASYH 68

QY 67 CDVTNEDGVKNADVNTSTYGLKIDIMFNSAGISDPNRPRIIDNEKADFERVLSVNVTVGVF 126

DB 69 CDVTNTEVENAVKFTVEKYGLDVLFLFSNAGVIEP-FVSILDLNLNLDRTIAINLRGTA 127

QY 127 LCMKHAARVMI-PARGSNIIISTASLSTMTGGSSSHAYCGSKHVALTLRNLAVALGQFGI 185

DB 128 AFIKHAARAMVEKGIKRSIVCTTSVAAETAGTAPHGYTTTSKHLLGLIKSASGGLGKYGI 187

QY 186 RYNCLSPFGLPTALGKFKFSIKNEEE-FENVINFAGNLKGPKNFVEDVANAAVYLASDEA 244

DB 188 RVNGVAPFGVATPL--VCNGFKMEPNVVEQNTSASANLKGIVLKHARHVAAPALFLASDES 245

QY 245 KYVSGHNLFTDGGFSV 260

DB 246 AYVSGQNLAVDGGYSV 261

RESULT 8

T02175

probable alcohol dehydrogenase At2g47130 [imported] - Arabidopsis thaliana

N:Alternate names: hypothetical protein F14M4.4

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001

C:Accession: T02175; D84911

R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; submitted to the EMBL Data Library, September 1998

A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.

A:Reference number: Z14609

A:Accession: T02175

A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-257 <ROU>
A:Cross-references: EMBL:AC004411; NID:g3522932; PID:g3522935
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84911
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <STO>
A:Cross-references: GB:AE002093; NID:g3522935; PIDN:AAC34217.1; GSPDB:GN00139
C:Gene: At2g47130; F14M4.4
A:Map position: 2
A:Introns: 5/2
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 40.7%; Score 567; DB 2; Length 257;
Best Local Similarity 45.4%; Pred. No. 1.4e-37;
Matches 118; Conservative 49; Mismatches 81; Indels 12; Gaps 5;

QY 10 RLEGKVALITGASGIGETAKLFSQHGAKVAITADVDLGHSHVVEAIGTSNHYHCDV 69
DB 5 RLDGKIALITGASGIGAEAVRLFTDHGAKVIVDPQDELGNQVAVSGKDKASFYRCDV 64

QY 70 TNEGCKNAVDNTYSTYKGLDIMPNSNAGISDPNRP-RIIDNEKADFERVLSVNVGTGVFLC 128
DB 65 TNEKEVENAVFTYKGYKGLDVLFSNAGVME--QPGSFLDLNLEQFDRTMVAVNRGAAG 122

QY 129 MKHAARVMI-PARSGNIISTASLSSTMGSSSHAYCGSKHAVLATRNLADELGOFGRV 187
DB 123 IKHAARVMI-PARSGNIISTASLSSTMGSSSHAYCGSKHAVLATRNLADELGOFGRV 182

QY 188 NCLSPGLPTALGKFKSGIKNEEFENFAGNLKGPKNVEDVANAALYLASDEAKYVSGH 244
DB 183 NGVAPYAVATAINS-----RDEETVRVMEYSAAATGILKGVVFKARHVAEALFLASDDS 237

QY 245 KYVSGHNLFDGFSVNSV 264
DB 238 AYVSGQNLAVDGGISVVKPI 257

RESULT 9
T06364
Probable short-chain alcohol dehydrogenase (EC 1.1.1.-) - tomato (fragment)
C:Species: Lycopersicon esculentum (tomato)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-May-2000
C:Accession: T06364
R:Jacobsen, S.E.; Olszewski, N.E.
Planta 198, 78-86, 1996
A>Title: Gibberellins regulate the abundance of RNAs with sequence similarity to protein
A:Reference number: Z15627; MUID:96158488; PMID:8580773
A:Accession: T06364
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-251 <JAC>
A:Cross-references: EMBL:U01801; NID:g717141; PIDN:AAB00109.1; PID:g717142
A:Experimental source: cultivar Moneymaker
C:Gene: GA03
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; oxidoreductase

Query Match 40.1%; Score 559; DB 2; Length 251;
Best Local Similarity 47.5%; Pred. No. 5.7e-37;
Matches 122; Conservative 46; Mismatches 81; Indels 8; Gaps 5;

QY 12 BGKVALITGASGIGETAKLFSQHGAKVAITADVDLGHSHVVEAIGTSNHYHCDVTN 71

Db 1 ECKVAIITGAASGIGESARLFEHGARVVVADIQDELGQKVVDISIGSKASYRHCDVTD 60

QY 72 EDGVKNVNDTSTYKGLDIMPNSNAGISDPNRPRIIDNEKADFERVLSVNVGTGVFLCMKH 131

Db 61 EKQVBETVAYAVEKYIGLIDIMPNSVGLT--NFCVLDMDVLAFFDETMANVR-IALAVKH 117

QY 132 AARVMIPAR-SGNIISTASLSSTMGSSSHAYCGSKHAVLATRNLADELGOFGRVNCV 190

Db 118 AAKVMVDKIRGSIICNASLEGILAGAASLAYTASHAVVGIKAARGLPHGRVNGV 177

QY 191 SPFGPLTALGKFKSGIKNEEFENFAGNLKGPKNVEDVANAALYLASDEAKYVSGH 250

Db 178 SPYGIATPLVTAKVGL-DAALLEAIYGNHGLKGVKLSMTMHAQSAFLASDESAYTSQ 236

QY 251 NLFIDGFSVNSVTKV 267

Db 237 NLAVDGGLS---SILKL 250

RESULT 10
T02176
Probable alcohol dehydrogenase At2g47120 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F14M4.5
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02176; C84911
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, September 1998
A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.
A:Reference number: Z14609
A:Accession: T02176
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-258 <ROU>
A:Cross-references: EMBL:AC004411; NID:g3522932; PID:g3522936
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84911
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <STO>
A:Cross-references: GB:AE002093; NID:g3522936; PIDN:AAC34218.1; GSPDB:GN00139
C:Gene: At2g47120; F14M4.5
A:Map position: 2
A:Introns: 5/2
A:Note: F14M4.5
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 39.4%; Score 549; DB 2; Length 258;
Best Local Similarity 44.5%; Pred. No. 3.3e-36;
Matches 113; Conservative 53; Mismatches 79; Indels 9; Gaps 5;

QY 10 RLEGKVALITGASGIGETAKLFSQHGAKVAITADVDLGHSHVVEAIGTSNHYHCDV 69

Db 5 RLEGKVALITGASGIGAEAVRLFTDHGAKVIVDPQDELGNQVAVSGKDKASFYRCDV 64

QY 70 TNEGCKNAVDNTYSTYKGLDIMPNSNAGISDPNRPRIIDNEKADFERVLSVNVGTGVFLCM 129

Db 65 TNETEVEDAVKFTYKGYKGLDVLFSNAGVLEP-LESFLDFDLFRDRAVNVNRGAAGAFI 123

QY 130 KHAARVMI-PARSGNIISTASLSSTMGSSSHAYCGSKHAVLATRNLADELGOFGRV 188

Db 124 KHAARVMI-PARSGNIISTASLSSTMGSSSHAYCGSKHAVLATRNLADELGOFGRV 182

QY 189 CLSPGLPTALCK--KFSGKNEEFENFAGNLKGPKNVEDVANAALYLASDEAKY 246

[illegible]

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: C70885
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, M.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A:Reference numbers: A70500; MUID:98295987; PMID:9634230
C:Accession: C70885
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-258 <COL>
A:Molecule type: DNA
C:Cross-references: GB:AL008883; GB:AL123456; NID:g3261490; PIDN:CAAL5519.1; PID:g261280
A:Experimental source: strain H37Rv
A:Gene: Rv2857c
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
F:11-188/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 28.9%; Score 402.5; DB 2; Length 258;
Best Local Similarity 35.8%; Pred. No. 1.5e-24;
Matches 91; Conservative 49; Mismatches 105; Indels 9; Gaps 4;

QY 8 ARRLGKVALITGCGASIGETTKAKFSQHCACAKVATADVDDELGHVSVEAIGTSNSTYIHC 67
DB 5 SORLAGRVAVITGGSGIGLAGRRMREGATIVYGDVDVEAGGAADDEL---SGLFVPT 61
QY 68 DVTNEDGVKNVAVNTVSTYCKLDMFNSNAGISDPNRPRIIDNEKADFERVLSVNVVTGVFL 127
DB 62 DVCDEDAVNGLFDCAAETTYGRIADFNAGISPPENLIENTELAAWQVDVNLKSYIL 121
QY 128 CMKHAARVMPARSNGIISTASSTMGSGSSH-AVCGSKHVALTRNLAVELGQFGIR 186
DB 122 CCRALRHMLVLAGKSVINVSFVAVGMSATSOISYASKGVLAMSRELGVQFARQGI 181
QY 187 VNCISPGFLPTALGKFSGKKEEFENVINF-AGNLGPKFENVEDVANAALYLASDEAK 245
DB 182 VNALCPGVNTPLQLQELFAKNPERAARMVHVLGRFAEP----DEIAAAVAFLASDDAS 237
QY 246 YVSGHNLFDGGFS 259
DB 238 FITASTFLVDGGIS 251

RESULT 15
AI0406
probable dehydrogenase YPO3351 [imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AI0406
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, N.; Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, N.
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
C:Accession: AI0406
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <KUR>
C:Cross-references: GB:AL590842; PIDN:CAC92581.1; PID:g15981278; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3351
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 28.6%; Score 398; DB 2; Length 256;
Best Local Similarity 37.4%; Pred. No. 3.4e-24;
Matches 99; Conservative 51; Mismatches 87; Indels 28; Gaps 9;

QY 5 TAFARRLEGKVALITGCGASIGETTKAKFSQHCACAKVATADVDDELGHVSVEAIGTSNSTY 64
DB 10 TAFS--LSGRVAITGGATGIGHAIELYLAARGVLMDCADNVA-EIAEQLDORDNAV 66

QY 65 IHCQVNTNEDGVKNVAVNTVSTYCKLDMFNSNAGIS--DPNRPRIIDNEKA---DFERVLS 119
DB 67 LHCQVSDSQSVRQVAQAIGAFAQGLDILYNSAGIAALDP-----AEKVRQDQWDRITD 119
QY 120 VNVTVGVELCMKHAARVMPARSNGIISTASLSTMGSGSSHAYCGSKHVALTRNLAVE 179
DB 120 INLKGVFLMCOEVGKHFIQHGKGIYNLASQAGVVALPNHLAYCASKFGVIGITVKVLA 179
QY 180 LQQFGIRVNCISPGFLPTALGKK-FSGIKNEEFENVINFAGNLKGP--KFNV-EDVANA 235
DB 180 WGPLDIOVNAISPTVVLTALGQKAWSQLAEDM-----CLKIPARRFAYPAEVAAC 230
QY 236 ALYLASDEAKYVSGHNLFDGGFSV 260
DB 231 ALFLASDAANMITGENELVIDGGYTI 255

Search completed: February 14, 2003, 09:20:37
Job time : 20 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:18:49 ; Search time 37 Seconds
(without alignments)
983.173 Million cell updates/sec

Title: US-09-673-918a-2
Perfect score: 1393
Sequence: 1 MQLRTAFARLEGRKVALITG.....IDGFGSVCNSIKVFQYDPS 273

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1393	100.0	273	21	AA1980.DAT
2	1388	99.6	272	21	AA1981.DAT
3	1364	97.9	277	21	AA1982.DAT
4	1233	88.5	276	21	AA1983.DAT
5	1233	88.5	276	23	AAO21494
6	1223	87.8	273	21	AA1984.DAT
7	1175	84.4	277	21	AA1985.DAT
8	658.5	47.3	285	23	ABB91329
9	654.5	47.0	285	21	AA1986.DAT
10	625.5	44.9	303	23	ABB92624

11	598.5	43.0	300	21	AA198620
12	598.5	43.0	343	21	AA198619
13	597.5	42.9	283	21	AA198619
14	597.5	42.9	283	23	ABB92823
15	588.5	42.2	336	16	AAW06488
16	588	42.2	269	23	ABB92448
17	578.5	41.5	261	21	AA198619
18	578.5	41.5	276	21	AA198619
19	575.5	41.3	264	23	ABB92145
20	575.5	41.3	274	21	AA198619
21	573.5	41.2	260	21	AA198619
22	573.5	41.2	260	23	ABB92475
23	569	40.8	280	23	ABB92449
24	568.5	40.8	286	21	AA198619
25	567	40.7	257	21	AA198619
26	567	40.7	257	23	ABB92144
27	566	40.6	259	23	ABB92476
28	565	40.6	257	21	AA198619
29	549.5	39.4	258	23	ABB92143
30	492.5	35.4	272	21	AA198619
31	492.5	35.4	275	21	AA198619
32	490.5	35.2	272	23	ABB92500
33	487.5	35.0	272	21	AA198619
34	432	31.0	253	22	AAU36249
35	403.5	29.0	159	21	AA198619
36	375.5	27.0	146	21	AA198619
37	370	26.6	186	21	AA198619
38	370	26.6	186	21	AA198619
39	370	26.6	186	21	AA198619
40	363	26.1	254	23	ABB48207
41	357	25.6	251	20	AA198619
42	349.5	25.1	262	22	AA198619
43	347.5	24.9	262	21	AA198619
44	346.5	24.9	253	23	ABB50010
45	343.5	24.7	267	22	AA198619

ALIGNMENTS

RESULT 1	
AA198619	AA198619 standard: Protein; 273 AA.
ID	AA198619 standard: Protein; 273 AA.
XX	AA198619 standard: Protein; 273 AA.
AC	AA198619 standard: Protein; 273 AA.
XX	AA198619 standard: Protein; 273 AA.
DT	06-APR-2000 (first entry)
XX	Secoisolaricresinol dehydrogenase protein clone DEHY133.
DE	Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;
KW	lignan; matairesinol; health-protecting lignan; phytoestrogen;
KW	enterolactone; enterodiol; neutriceutical; dietary supplement;
KW	(-)-trachelogenin; antiviral; (-)-podophyllotoxin.
XX	Forsythia intermedia.
OS	Forsythia intermedia.
XX	WO9955846-A1.
PN	WO9955846-A1.
XX	04-NOV-1999.
PD	04-NOV-1999.
XX	23-APR-1999; 99WO-US08975.
PF	23-APR-1999; 99WO-US08975.
XX	24-APR-1998; 98US-0082977.
PR	(UNIW) UNIV WASHINGTON STATE RES FOUND.
XX	(UNIW) UNIV WASHINGTON STATE RES FOUND.
PA	(UNIW) UNIV WASHINGTON STATE RES FOUND.
XX	Xia Z, Costa MA, Davin LB, Lewis NG;
PI	Xia Z, Costa MA, Davin LB, Lewis NG;
XX	WPI; 2000-126356/11.
DR	N-PSDB; AA45730.
XX	New nucleic acid molecule encoding an enzyme involved in lignan
PT	New nucleic acid molecule encoding an enzyme involved in lignan

PT biosynthetic pathway, useful for producing large amounts of lignans -
 XX Claim 8; Page 46-47; 66pp; English.
 PS The present sequence represents a secoisolariciresinol dehydrogenase
 XX protein. The enzyme is involved in the lignan biosynthetic pathway.
 CC The secoisolariciresinol dehydrogenase proteins have a molecular
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
 CC nucleic acids are used for the recombinant expression of the enzymes.
 CC It is also used to obtain expression or enhanced expression of
 CC secoisolariciresinol dehydrogenase in plants or to alter lignan
 CC biosynthesis. The enzyme is used for production of the pharmacologically
 CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase
 CC proteins and nucleic acids can be utilized to: elevate or otherwise
 CC alter the levels of health-protecting lignans, including phytoestrogens
 CC such as enterolactone and enterodiol, in plant species, including
 CC vegetables, grains and fruits and to food items incorporating material
 CC derived from such genetically altered plants; genetically alter plant
 CC species to provide an abundant, natural supply of lignans useful for
 CC a variety of purposes, for example as nutraceuticals and dietary
 CC supplements; to genetically alter living organisms to produce an
 CC abundant supply of optically pure lignans having desirable biological
 CC properties, for example (-)-trachelogenin which possesses antiviral
 CC properties, and (-)-podophyllotoxin.
 XX Sequence 273 AA;
 SQ
 Query Match 100.0%; Score 1393; DB 21; Length 273;
 Best Local Similarity 100.0%; Pred. No. 1.8e-132;
 Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQLTAFARLEGKVALITGGASGIGETTAKLFSQHGAKVAIADVDLGHVSVEAIGTS 60
 DB 1 MQLTAFARLEGKVALITGGASGIGETTAKLFSQHGAKVAIADVDLGHVSVEAIGTS 60
 QY 61 NSTYIHCVDVTNEDGVKNVADNTVSTYKLDIMFNSNAGISDPNRPRIIDNEKADFERVLSV 120
 DB 61 NSTYIHCVDVTNEDGVKNVADNTVSTYKLDIMFNSNAGISDPNRPRIIDNEKADFERVLSV 120
 QY 121 NVTGVFLCMKHAARVMPARSGNIISTASLSMTGGGSHAYCGSKHAVLATRNLADEL 180
 DB 121 NVTGVFLCMKHAARVMPARSGNIISTASLSMTGGGSHAYCGSKHAVLATRNLADEL 180
 QY 181 QGQIRVNCCLSPFGLPTALGKFGSGIKNEEFENVINFAGNLKGPKNVEDVANAALYLA 240
 DB 181 QGQIRVNCCLSPFGLPTALGKFGSGIKNEEFENVINFAGNLKGPKNVEDVANAALYLA 240
 QY 241 SDEAKYVSGHNLFDIDGGFVSCNSVIKVFQYDPS 273
 DB 241 SDEAKYVSGHNLFDIDGGFVSCNSVIKVFQYDPS 273
 RESULT 2
 ID AAY54420
 XX AAY54420 standard; Protein; 272 AA.
 XX AC AAY54420;
 XX DT 06-APR-2000 (first entry)
 XX DE Secoisolariciresinol dehydrogenase protein clone DEHV130.
 XX KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
 XX RN lignan; matairesinol; health-protecting lignan; phytoestrogen;
 XX KW enterolactone; enterodiol; nutraceutical; dietary supplement;
 XX KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin.
 XX OS Forsythia intermedia.
 XX FN W0995846-A1.
 XX PD 04-NOV-1999.

XX 23-APR-1999; 99WO-US08975.
 PF 24-APR-1998; 98US-0082977.
 PR (UNIW) UNIV WASHINGTON STATE RES FOUND.
 XX Xia Z, Costa MA, Davin LB, Lewis NG;
 DR WPI: 2000-126356/11.
 DR N-PSDB; AAZ45743.
 PT New nucleic acid molecule encoding an enzyme involved in lignan
 PT biosynthetic pathway, useful for producing large amounts of lignans -
 XX Example 2: Page 61-63; 66pp; English.
 XX The present sequence represents a secoisolariciresinol dehydrogenase
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.
 CC The secoisolariciresinol dehydrogenase proteins have a molecular
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
 CC nucleic acids are used for the recombinant expression of the enzymes.
 CC It is also used to obtain expression or enhanced expression of
 CC secoisolariciresinol dehydrogenase in plants or to alter lignan
 CC biosynthesis. The enzyme is used for production of the pharmacologically
 CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase
 CC proteins and nucleic acids can be utilized to: elevate or otherwise
 CC alter the levels of health-protecting lignans, including phytoestrogens
 CC such as enterolactone and enterodiol, in plant species, including
 CC vegetables, grains and fruits and to food items incorporating material
 CC derived from such genetically altered plants; genetically alter plant
 CC species to provide an abundant, natural supply of lignans useful for
 CC a variety of purposes, for example as nutraceuticals and dietary
 CC supplements; to genetically alter living organisms to produce an
 CC abundant supply of optically pure lignans having desirable biological
 CC properties, for example (-)-trachelogenin which possesses antiviral
 CC properties, and (-)-podophyllotoxin.
 XX Sequence 272 AA;
 SQ
 Query Match 99.6%; Score 1388; DB 21; Length 272;
 Best Local Similarity 100.0%; Pred. No. 5.7e-132;
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 QLRTAFAARLEGKVALITGGASGIGETTAKLFSQHGAKVAIADVDLGHVSVEAIGTSN 61
 DB 1 QLRTAFAARLEGKVALITGGASGIGETTAKLFSQHGAKVAIADVDLGHVSVEAIGTSN 60
 QY 62 STYIHCVDVTNEDGVKNVADNTVSTYKLDIMFNSNAGISDPNRPRIIDNEKADFERVLSV 121
 DB 61 STYIHCVDVTNEDGVKNVADNTVSTYKLDIMFNSNAGISDPNRPRIIDNEKADFERVLSV 120
 QY 122 VTGVFLCMKHAARVMPARSGNIISTASLSMTGGGSHAYCGSKHAVLATRNLADEL 181
 DB 121 VTGVFLCMKHAARVMPARSGNIISTASLSMTGGGSHAYCGSKHAVLATRNLADEL 180
 QY 182 QGIRVNCCLSPFGLPTALGKFGSGIKNEEFENVINFAGNLKGPKNVEDVANAALYLA 241
 DB 181 QGIRVNCCLSPFGLPTALGKFGSGIKNEEFENVINFAGNLKGPKNVEDVANAALYLA 240
 QY 242 DEAKYVSGHNLFDIDGGFVSCNSVIKVFQYDPS 273
 DB 241 DEAKYVSGHNLFDIDGGFVSCNSVIKVFQYDPS 272
 RESULT 3
 ID AAY54413
 XX AAY54413 standard; Protein; 277 AA.
 XX AC AAY54413;
 XX DT 06-APR-2000 (first entry)

Db 5 QLRFAFARLEGGKVALITGGASGVGEVTAQLFSQHGAKVAIADVQDELGHVSVEAIGLSN 64
 QY 62 STYTHCDVTNEDGVKNVADNTVSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSVN 121
 Db 65 STYTHCDVTNEDGVKNVADNTVSTYKGLDIFMSNAGISDPYKPRVIDNEKADFERVLSVN 124
 QY 122 VTGFLCMKHAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAYLALTRNLAVELG 181
 Db 125 VTGFLCMKHAARVMIPARSGCIISTASLSSTMGSGSHAYCGSKHAYLALTRNLAVELG 184
 QY 182 QFGRVNCCLSPFGLPTALGKFKSGIKNEEFENVINPAGNLKGPKNFVEDVANAALYLAS 241
 Db 185 QFGRVNCCLSPFGLPTALGKFKSGIKNEEFENVINPAGNLKGPKNFVEDVANAALYLAS 244
 QY 242 DEAKYVSGHNLFDIDGFSVCNSVVKVQYFDPDS 273
 Db 245 DEAQVSGQNLFDIDGFSVCNSAIKLFOYFDPDS 276

RESULT 5
 AAO21494
 ID AAO21494 standard; Protein; 276 AA.
 XX
 AC AAO21494;
 XX
 DT 15-AUG-2002 (first entry)
 DE Secoisolariciresinol dehydrogenase protein.
 XX
 KW Guaiacyl (G)-lignan; monocotyledon plant; rice; food additive;
 KW seed-specific transcriptional regulatory region; dehydrogenase; enzyme.
 XX
 OS Unidentified.
 XX
 PN WO200220548-A1.
 XX
 PD 14-MAR-2002.
 PF 04-SEP-2001; 2001WO-0527500.
 XX
 PR 07-SEP-2000; 2000US-230632P.
 XX
 PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
 PA (PHYT-) APPLIED PHYTOLOGICS INC.
 XX
 PI Lewis NG, Davin LB, Huang N;
 XX
 DR WPI; 2002-425767/45.
 DR N-PSDB; AAL38432.
 XX
 PT Increasing guaiacyl-lignan content in monocotyledon plants, by
 PT transforming plant with chimeric gene construct having seed-specific
 PT transcriptional regulator linked to gene encoding protein involved in
 PT G-lignan formation
 XX
 PS Claim 2; Fig 10; 136pp; English.
 XX
 CC The invention relates to a method for increasing the guaiacyl (G)-lignan
 CC content in seeds of a monocotyledon plant, comprising selecting at least
 CC one protein or enzyme integral to the pathway leading to G-lignan
 CC formation, stably transforming a monocotyledon plant with chimeric gene
 CC (CG) constructs having a seed-specific transcriptional regulatory region
 CC operably linked to a nucleic acid sequence encoding the enzyme. The
 CC method of the invention is useful for stably transforming a
 CC monocotyledonous plant (e.g. rice) with CG constructs resulting in
 CC increased expression of the genes encoded by CG constructs. The G-lignan
 CC enriched seed composition is useful as a food additive. This
 CC sequence represents the secoisolariciresinol dehydrogenase protein
 CC relating to the invention.
 XX
 SQ Sequence 276 AA;
 Query Match 88.5%; Score 1233; DB 23; Length 276;

Best Local Similarity 88.2%; Pred. No. 2.8e-116;
 Matches 240; Conservative 14; Mismatches 18; Indels 0; Gaps 0;

QY 2 QLRFAFARLEGGKVALITGGASGIGETTAQLFSQHGAKVAIADVQDELGHVSVEAIGTSN 61
 Db 5 QLRFAFARLEGGKVALITGGASGVGEVTAQLFSQHGAKVAIADVQDELGHVSVEAIGLSN 64
 QY 62 STYTHCDVTNEDGVKNVADNTVSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSVN 121
 Db 65 STYTHCDVTNEDGVKNVADNTVSTYKGLDIFMSNAGISDPYKPRVIDNEKADFERVLSVN 124
 QY 122 VTGFLCMKHAARVMIPARSGNIISTASLSSTMGSGSHAYCGSKHAYLALTRNLAVELG 181
 Db 125 VTGFLCMKHAARVMIPARSGCIISTASLSSTMGSGSHAYCGSKHAYLALTRNLAVELG 184
 QY 182 QFGRVNCCLSPFGLPTALGKFKSGIKNEEFENVINPAGNLKGPKNFVEDVANAALYLAS 241
 Db 185 QFGRVNCCLSPFGLPTALGKFKSGIKNEEFENVINPAGNLKGPKNFVEDVANAALYLAS 244
 QY 242 DEAKYVSGHNLFDIDGFSVCNSVVKVQYFDPDS 273
 Db 245 DEAQVSGQNLFDIDGFSVCNSAIKLFOYFDPDS 276

RESULT 6
 AAV54414
 ID AAV54414 standard; Protein; 273 AA.
 XX
 AC AAV54414;
 XX
 DT 06-APR-2000 (first entry)
 DE Secoisolariciresinol dehydrogenase protein clone SMDEHY431.
 XX
 KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
 KW lignan; matairesinol; health-protecting lignan; phytoestrogen;
 KW enterolactone; enterodiol; neutriceutical; dietary supplement;
 KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin.
 XX
 OS Forsythia intermedia.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 120
 FT Misc-difference 120 /note= "encoded by GTN"
 FT Misc-difference 122
 FT Misc-difference 122 /note= "encoded by GTN"
 XX
 PN WO9955846-A1.
 XX
 PD 04-NOV-1999.
 XX
 PF 23-APR-1999; 99WO-US08975.
 XX
 PR 24-APR-1998; 98US-0082977.
 XX
 PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
 XX
 PI Xia Z, Costa MA, Davin LB, Lewis NG;
 XX
 DR WPI; 2000-126356/11.
 DR N-PSDB; AAZ45732.
 XX
 PT New nucleic acid molecule encoding an enzyme involved in lignan
 PT biosynthetic pathway, useful for producing large amounts of lignans
 XX
 PS Claim 8; Page 51-52; 66pp; English.
 XX
 CC The present sequence represents a secoisolariciresinol dehydrogenase
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.
 CC The secoisolariciresinol dehydrogenase proteins have a molecular
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
 CC nucleic acids are used for the recombinant expression of the enzymes.


```
XX 28-AUG-2001; 2001WO-EP09892.
XX 28-AUG-2001; 2001WO-EP09892.
XX (FARB ) BAYER AG.
XX Tietjen K, Weidler M;
XX WPT; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plants with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 540; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 285 AA;
SQ
Query Match 47.3%; Score 658.5; DB 23; Length 285;
Best Local Similarity 50.9%; Pred No. 4.3e-58;
Matches 139; Conservative 43; Mismatches 78; Indels 13; Gaps 5;
QY 8 ARRLGKVALITGASGIGETRAKLFQSGHAKVAIADVDGLHGVPEAI--GTSNST-- 63
Db :||| ||||| ||||| :|| :||| ||| :||| ||| :|| :|||
QY 64 YHCDVTNEDGVKNVDNTVSYGKLDIMFSNAGISDPNRPRIIDNEKADPERVLSVNV 123
Db :||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 75 FHGDVRVEDDIDNAVDFAVKNFTGLDILINAGLCGAPCPDIRNYSLEFEMTFDYNV 134
Db :||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 124 GVFLCMKHAARVIMPARSGNIISTASLSMTGGSSHAYCGSKHAVLALTRNLAVELGQ 183
Db :||| ||||| ||||| :||| :||| :||| ||||| ||||| :|||
QY 135 GAFLSMKHAARVIMPEKGSIVSLCSVGVGVGVGPHSYGSKHAVLGLTRVAALGQH 194
Db :||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 184 GIRVNCLSFGLPTALGKFKFSIGIKNEEPEFNI----NFA---GNLKGPFNVEDVANA 236
Db :||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 195 GIRVNCVSPYAVATKLA--LAHLPEERTEDEAFVGRFAAANALKGVELTVDDVANAV 252
Db :||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 237 IYLASDEAKYVSGHNLFDGGFSVCNSVIKVFQ 269
Db :||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 253 LFLASDDSRVISGNDLMIDGGFTCTNHSFKVFR 285
Db :||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
RESULT 9
AAG2552A
ID AAG2552A standard; Protein; 285 AA.
XX
AC AAG2552A;
XX
DI 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 29625.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
```

```
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134371.
XX 18-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 21-MAY-1999; 99US-0135124.
XX 24-MAY-1999; 99US-0135353.
XX 25-MAY-1999; 99US-0135629.
XX 27-MAY-1999; 99US-0136021.
XX 28-MAY-1999; 99US-0136392.
XX 01-JUN-1999; 99US-0136782.
XX 03-JUN-1999; 99US-0137222.
XX 04-JUN-1999; 99US-0137528.
XX 07-JUN-1999; 99US-0137502.
XX 08-JUN-1999; 99US-0137724.
XX 10-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 14-JUN-1999; 99US-0138847.
XX 16-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 17-JUN-1999; 99US-0139453.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 21-JUN-1999; 99US-0139763.
XX 22-JUN-1999; 99US-0139817.
XX 23-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 24-JUN-1999; 99US-0140354.
XX 28-JUN-1999; 99US-0140695.
XX 29-JUN-1999; 99US-0140823.
XX 30-JUN-1999; 99US-0140991.
XX 01-JUL-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 02-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
```



```
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
```

```
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 47.0%; Score 654.5; DB 21; Length 285;

Best Local Similarity 50.9%; Pred. No. 1.1e-57; Mismatches 42; Indels 13; Gaps 5;

```
Matches 139; Conservative 42; Mismatches 79; Indels 13; Gaps 5;

QY 8 ARRECKVALITGGAGIGETTAKLFQHQAKVAIADYQDELGHVSVEAI--GTSNST-- 63
   ::|| |||||::||:: ||::|||::|||::|||::|||::|||::|||
Db 15 SQRLLGKVALITGGATGICESIVLFHKGAKVCIVLDQDLGGGVCKSLRGESKETAF 74
   :|| || ||::||| ||::|||::|||::|||::|||::|||::|||
QY 64 YIHCDVTNEDGVKNVNDVTSTYTKLDMFSNAGISDPNRPRIIDNEKADPERVLSVNT 123
   :|| || ||::||| ||::|||::|||::|||::|||::|||::|||
Db 75 FTHGDVRVEDDISNAVDFAVKNFTLDILINNAGLCGAPCPDIRNYSLSEFEFETDVNVK 134
   :|| || ||::||| ||::|||::|||::|||::|||::|||::|||
QY 124 GVFLCMKHAARVMIPARSGNIISTASLSTMGSSGHAYCGSKHAVLALTRNLAVELQOF 183
   || || |||||::|||::|||::|||::|||::|||::|||::|||
Db 135 GAFLSMKHAARVMIPKKGSIIVSLCVSGVGVGPHSVGSKHAVLGLTRSVAAELGOH 194
   |||||::|||::|||::|||::|||::|||::|||::|||
QY 184 GIRVNCVSPFGLPTALGKKFGIKNEEEFENVI----NFA----GNLKGPKFNVEDVANAA 236
   |||||::|||::|||::|||::|||::|||::|||::|||
Db 195 GIRVNCVSPYAVATKLA--LAHLPEERTEDAFVGRNFAANANLKGVELTVGDVANAV 252
   :|||::|||::|||::|||::|||::|||::|||::|||
QY 237 LYLASDEAKYVSGHNLFDGGFSCVNSVIKVFQ 269
   :|||::|||::|||::|||::|||::|||::|||::|||
Db 253 LFLASDDSRYSIGDNLMDIDGGFTCTNHSFKVPR 285
   :|||::|||::|||::|||::|||::|||::|||::|||
```

RESULT 10

ABB92624

ID ABB92624 standard; Protein; 303 AA.

XX ABB92624;

XX ABB92624;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1835.

PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149903.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 43.0%; Score 598.5; DB 21; Length 300;		
Best Local Similarity 46.8%; Pred. No. 5.6e-52;		
Matches 126; Conservative 46; Mismatches 86; Indels 11; Gaps 4;		
Qy	8	ARRLEKVALITGGASGIGETTAKLFQSHQCAKVAADVQDELGHSVVEAIGTNSNTHC 67
Db	32	SRKLEKVALITGGASGIGKATAGFISHGAKVIIADIQPQIGRETEQELGFS-CAYFPC 90
Qy	68	DVTNEDGVKNVNDTVSTYKLDIMFSNAGISDPNRPRIIDNEKADFERVLSVNVTVGL 127
Db	91	DVTKESDIANAVDAVSLHTKLDIMYNNAGIECKTPPSIVDLNVDKVINTVRGVMA 150
Qy	128	CMKHAARVMIPARSGNIITSTASLSMTGGGSHAYCGSKHAVLALTRNLAVELGQFGRV 187
Db	151	GIKHAARVMIPRNSGSIICAGSVTGMGLAQHTYSVKSASVIGIVRSTASELCKHRIV 210
Qy	188	NCLSPFGLPTA-----LGKKFSGINEEFENVINFAGNLKGPKNVEVDANAALYASD 242
Db	211	NCISPFAITTSFVMDMRQIYFV-DSRLIQIVOSTGVINGEVCPTDVAANAAYLASD 269
Qy	243	EAKYVSGHNLFIIDGGFSVCNSVIKVFQYP 271
Db	270	DSKYVNGHNLVVDGGF-----TTVKTLDFP 294
RESULT 12		

AAG18619
ID AAG18619 standard; Protein; 343 AA.
XX
AC AAG18619;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20100.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132487.
PR 14-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142054.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139751.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145313.
 PR 27-JUL-1999; 99US-0145318.
 PR 27-JUL-1999; 99US-0145319.
 PR 28-JUL-1999; 99US-0145351.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158569.
 PR 13-OCT-1999; 99US-0159233.
 PR 13-OCT-1999; 99US-0159234.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 597.5; DB 21; Length 283;
 Best Local Similarity 47.0%; Pred. No. 6.4e-52;
 Matches 126; Conservative 45; Mismatches 86; Indels 11; Gaps 4;
 OY 9 RLLEKVALITGGASGIGETAKLFSGHGAKVATADVDQLGHVSVEATGTSNYSYIHCD 68
 Db 16 RKLEKVALITGGASGIGKATAGKFIHGAKVIIADIQIGRETEQELGPS-CAFFCD 74

```
QY 69 VTNEGDKNAVDNTVSTYKGLDIFESNAGISDPNRPRIIDNEKADFERVLSVNTGVFLC 128
  ||| : |||| || : ||||| : ||||| ||| : ||| : ||| : |||
Db 75 VTKESDIANAVDFAVSLHTKLDIMYNAGIPCKTPPSIVDLNLNFKVINTNVRGVMA 134
  ||| : |||| || : ||||| : ||||| ||| : ||| : ||| : |||
QY 129 MKHAARVMIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGIRVN 188
  : ||||| ||||| ||||| : ||| : ||| : ||| : ||| : ||| : |||
Db 135 IKAARVMIPRNSGSIICAGSVTGMGGLAQHTYSVSKSAVIGIVRSTASELCKHRIRVN 194
  : ||||| ||||| ||||| : ||| : ||| : ||| : ||| : ||| : |||
QY 189 CLSPFGLPTA-----LGKKFSGIKNEEFENVINFAGNLKGPKFNVEDYANAALYLASDE 243
  : |||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 195 CISPPFAITTSFYVMDENMRQIYPGV-DDSRLIQIVQSTGVNLNGEVCPTDVANAAYLASDD 253
  : |||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 AKYVSGHNLFIDGGFSCVNSVTKVQYFP 271
  : ||||| ||||| : |||| : ||| : ||| : ||| : ||| : ||| : |||
Db 254 SKYVNGHNLVVDGGF-----TTVKTLDFP 277
  : ||||| ||||| : |||| : ||| : ||| : ||| : ||| : ||| : |||

RESULT 14
AAW06488
ID ABB92823 standard; Protein; 283 AA.
XX
AC ABB92823;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2034.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
WPI; 2002-269010/31.
XX
DR
XX
Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 2034; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 283 AA;

Query Match 42.9%; Score 597.5; DB 23; Length 283;
Best Local Similarity 47.0%; Pred. No. 6.4e-52;
Matches 126; Conservative 45; Mismatches 86; Indels 11; Gaps 4;

QY 9 RLLEKVALITGGASGIGETTAKLFQSHGAKVAIADVQDELGHVSVEAIGTSNHYIHD 68
  : ||||| ||||| ||||| : ||| : ||||| ||||| : ||| : ||| : |||
Db 16 RLLEKVALITGGASGIGATAGKTFISHGAKVLIADIQPIGRETEQELGPS-CAFFPCD 74
  : ||||| ||||| ||||| : ||| : ||||| ||||| : ||| : ||| : |||
QY 69 VTNEGDKNAVDNTVSTYKGLDIFESNAGISDPNRPRIIDNEKADFERVLSVNTGVFLC 128
```

```
Db 75 VTKESDIANAVDFAVSLHTKLDIMYNAGIPCKTPPSIVDLNLNFKVINTNVRGVMA 134
  ||| : |||| || : ||||| : ||||| ||| : ||| : ||| : |||
QY 129 MKHAARVMIPARSGNIISTASLSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFGIRVN 188
  : ||||| ||||| ||||| : ||| : ||| : ||| : ||| : ||| : |||
Db 135 IKAARVMIPRNSGSIICAGSVTGMGGLAQHTYSVSKSAVIGIVRSTASELCKHRIRVN 194
  : ||||| ||||| ||||| : ||| : ||| : ||| : ||| : ||| : |||
QY 189 CLSPFGLPTA-----LGKKFSGIKNEEFENVINFAGNLKGPKFNVEDYANAALYLASDE 243
  : |||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 195 CISPPFAITTSFYVMDENMRQIYPGV-DDSRLIQIVQSTGVNLNGEVCPTDVANAAYLASDD 253
  : |||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 AKYVSGHNLFIDGGFSCVNSVTKVQYFP 271
  : ||||| ||||| : |||| : ||| : ||| : ||| : ||| : ||| : |||
Db 254 SKYVNGHNLVVDGGF-----TTVKTLDFP 277
  : ||||| ||||| : |||| : ||| : ||| : ||| : ||| : ||| : |||

RESULT 15
AAW06488
ID AAW06488 standard; Protein; 336 AA.
XX
AC AAW06488;
XX
DT 10-FEB-1997 (first entry)
XX
DE Maize Ts2 sequence.
XX
KW Tasselseed 2; Ts2; maize; H2 clone; rice; floral tissue; male; female;
KW bacterial hydroxysteroid dehydrogenase; sexual development; flower.
XX
OS Zea mays.
XX
PN WO9505732-A1.
XX
PD 02-MAR-1995.
XX
PF 23-AUG-1994; 94WO-US09513.
XX
PR 23-AUG-1993; 93US-0110690.
XX
PA (UYVA ) UNIV YALE.
XX
PI Dellaporta SL;
XX
WPI; 1995-106591/14.
XX
N-PSDB; AAT45060.
XX
Tasselseed 2 gene and promoter and expression systems for
PT transforming plants - for controlling sexual development to produce
PT either predominantly male or female flowers, to simplify
PT cross-breeding and production of hybrid seeds.
XX
PS Claim 2; Fig 6; 44pp; English.
XX
CC This sequence represents the tasselseed 2 (Ts2) protein derived from
CC maize. This sequence is encoded by the H2 clone. The Ts2 protein in
CC maize is very similar to Ts2 protein found in rice, and both these
CC sequences show considerable homology to bacterial hydroxysteroid
CC dehydrogenases. Ts2 has a mol. wt. of 35 kD and a predicted pI of
CC 6.7. The Ts2 coding sequence and its promoter, may be used to control
CC sexual development of floral tissue and to provide exclusively male
CC or female flowers.
XX
SQ Sequence 336 AA;

Query Match 42.2%; Score 588.5; DB 16; Length 336;
Best Local Similarity 42.2%; Pred. No. 6.8e-51;
Matches 122; Conservative 50; Mismatches 88; Indels 29; Gaps 4;

QY 5 TAFARLECKVALITGGASGIGETTAKLFQSHGAKVAIADVQDELGHVSVEAIGTSNSTY 64
  : ||||| ||||| ||||| : ||| : ||||| ||||| : ||| : ||| : |||
Db 47 TPMPLKDGKVAIVTGGARGIGEAIVRLFAGKARVVIADIDDAAGALASALGPQVS-F 105
  : ||||| ||||| ||||| : ||| : ||||| ||||| : ||| : ||| : |||
QY 65 IHCDDVTNEDGVKNAVDNTVSTY-GKLDIFESNAGI---SDPNRPRIIDNEKADFERVLSV 120
```

Db 106 VRCDSVEDDVRRAVDWALSRHGRGLDVYCNAGVLGRQTRAARSILLSFDDAAEFDRVLRV 165
Qy 121 NVTGVFLCMKHAARVMIIPARSGNIIISTASLSTMGSSSHAYCGSKHAVLALTRNLAVEL 180
Db 166 NALGAALGMKHAARAMAPRAGSIVSVASVAVLGGLGPHAYTASKHAIVCLTKNAACEL 225
Qy 181 GQFGIRVNCLSPEGLPTA-----LGKKFSGIKNEEEFENV 216
Db 226 RAHGRVNCVSPFEGVATPMLINAWROGHDDATADADRDLDLDVTVPSDQEVKMEEVV 285
Qy 217 NFAGNLKGPKNVEDVANAAALYLASDEAKYVSGHNLFTIDGGFSVCNSVI 265
Db 286 RGLATLKGPTLRPRDIAEAVLFIAASDEARYISGHNLVVDGGVTTSRNLI 334

Search completed: February 14, 2003, 09:19:36
Job time : 38 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2003, 09:20:43 : Search time 13 Seconds
(without alignments)
536.526 Million cell updates/sec

Title: us-09-673-918A-2

Perfect score: 1393

Sequence: 1 MQLRTAFARLEKGVKALITG.....IDGFGSVNVIKVPQPS 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/FCI_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/FCIUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1233	88.5	276	9	US-09-944-160-24
2	432	31.0	253	10	US-09-815-242-11842
3	369	26.5	256	10	US-09-954-314-14
4	349.5	25.1	262	10	US-09-802-853-6
5	343.5	24.7	261	9	US-09-773-748-1
6	341.5	24.5	267	10	US-09-940-037A-29
7	332	23.8	263	10	US-09-815-242-14092
8	331	23.8	271	10	US-09-815-242-13387
9	331	23.8	272	10	US-09-815-242-13679
10	328.5	23.6	261	10	US-09-815-242-10846
11	328.5	23.6	272	10	US-09-815-242-12688
12	326	23.4	245	10	US-09-796-089-9
13	325	23.3	261	9	US-09-971-536-74
14	321.5	23.1	251	10	US-09-815-242-5689
15	321.5	23.1	270	9	US-09-978-295A-468
16	321.5	23.1	270	9	US-09-978-697-468
17	321.5	23.1	270	9	US-09-978-192A-468
18	321.5	23.1	270	9	US-09-999-832A-468
19	321.5	23.1	270	9	US-09-978-189-468

20	319.5	22.9	254	9	US-09-978-758-2	Sequence 2, Appli
21	319.5	22.9	258	9	US-10-156-055A-1	Sequence 1, Appli
22	318.5	22.9	246	10	US-09-739-254-143	Sequence 143, App
23	318.5	22.9	246	10	US-09-904-615-143	Sequence 143, App
24	318.5	22.9	253	10	US-09-931-186-15	Sequence 15, Appl
25	316.5	22.7	244	10	US-09-815-242-10126	Sequence 10126, A
26	310	22.3	285	10	US-09-815-242-10241	Sequence 10241, A
27	303	21.8	255	10	US-09-931-186-14	Sequence 14, Appl
28	301.5	21.6	246	10	US-09-815-242-5461	Sequence 5461, Ap
29	301.5	21.6	246	10	US-09-815-242-12123	Sequence 12123, A
30	301.5	21.6	246	10	US-09-815-242-12803	Sequence 12803, A
31	301.5	21.6	246	10	US-09-815-242-13100	Sequence 13100, A
32	296.5	21.3	262	10	US-09-815-242-10597	Sequence 10597, A
33	294.5	21.1	286	10	US-09-815-242-11860	Sequence 11860, A
34	288.5	20.7	274	9	US-09-976-059-17	Sequence 17, Appl
35	288	20.7	254	10	US-09-741-669-411	Sequence 411, App
36	287.5	20.6	242	10	US-09-815-242-10974	Sequence 10974, A
37	287	20.6	247	9	US-09-479-040-9	Sequence 9, Appli
38	283.5	20.4	245	10	US-09-815-242-10921	Sequence 10921, A
39	281	20.2	203	10	US-09-823-901-7	Sequence 7, Appli
40	277	19.9	258	9	US-09-738-626-6436	Sequence 6436, Ap
41	275	19.7	221	10	US-09-764-853-553	Sequence 553, App
42	274.5	19.7	206	9	US-09-997-816-4	Sequence 4, Appli
43	274.5	19.7	266	9	US-09-981-353-40	Sequence 40, Appl
44	274.5	19.7	266	10	US-09-931-186-18	Sequence 18, Appl
45	274	19.7	251	10	US-09-922-501-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-944-160-24
; Sequence 24, Application US/09944160
; Patent No. US20020174452A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, No. US20020174452Alman
; APPLICANT: Davin, Laurence
; APPLICANT: .. Huang, Ning
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan
; TITLE OF INVENTION: Content
; FILE REFERENCE: WSURL17983
; CURRENT APPLICATION NUMBER: US/09/944,160
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 60/230,632
; PRIOR FILING DATE: 2000-09-07
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: secoisolariciresinol dehydrogenase amino acid
; OTHER INFORMATION: sequence from plasmid PAP1249
US-09-944-160-24

Query Match	88.5%	Score	1233	DB	9	Length	276
Best Local Similarity	88.2%	Pred. No.	3.8e+108	Mismatches	14	Indels	0
Matches	240	Conservative	14	Gaps	0		
QY	2	QLRTAFARLEKGVKALITGASGIGETTAKLFSQHKAKVAIADVDQDELGHVSVEAIGTSN	61				
Db	5	QLRTAFARLEKGVKALITGASGIGETTAKLFSQHKAKVAIADVDQDELGHVSVEAIGLSN	64				
QY	62	STYIHCDVTNEDGVKNVADNTVSTYTKLDMFNSNAGISDPNRPRIIDNEKADFVRLSVN	121				
Db	65	STYIHCDVTNEDGVKNVADNTVSTYTKLDMFNNAGISDPYKPRVIDNEKADFVRLSVN	124				
QY	122	VTCVFLCMKHAARVMTIPARSGNIISTASLSSTMGSSSHAYCGSKHAVIALTRNLAVELG	181				
Db	125	VTCVFLCMKHAARVMTIPARSGNIISTASLSSTMGSSSHAYCGSKHAVIAGLTRNLAVELG	184				


```

Best Local Similarity 31.2%; Pred. No. 2.2e-25;
Matches 83; Conservative 55; Mismatches 107; Indels 21; Gaps 6;

QY 8 ARLECKVALITGGASGIGETAKLFSSQHGAKVAIADVQDEL---GHSVVEAIGTSNSTY 64
Db 2 SKRFNGKCLVGTGAGNIGLALRLAEETGTAIALLDMNREALAKAEASVRKGVVEARSY 61
QY 65 IHCVDVTNEDGVKNVNDVTSTYKGLDIFMSNAGISDPNRPRIIDNEKADFERVLSVNVVTG 124
Db 62 V-CDVTSEEAVTGVDSVVRDEFKIDFENNAGYQGAFAF-VQDYPSDDFAFVLTINVTG 119
QY 125 VFLCMKHAARVMIPARSGNIISTASLSSTMTGGSSHAYCGSKHAVLALTRNLAVELGQFG 184
Db 120 AFHLKAVSRQMITQNYGRIVNTASAGYKPPNMAAYGASKGAIITALTETAAOLDLAPYN 179
QY 185 IRVNCCLSP-----FGLPTALGKKFSGIKNEEFENFVINFAGNLKGPKE-NVED 231
Db 180 IRVNAISPGYMGPGFWMERQVLEQAKVGQSYFTDPKVVAAQMI---GSPMRRYGDINE 236
QY 232 VANAALYLASDEAKYVSGHNLFDGG 257
Db 237 IPGVVAFLLGDDSSFTGYNLPIAGG 262

RESULT 5
US-09-773-748-1
; Sequence 1, Application US/09773748
; Publication No. US20020187537A1
; GENERAL INFORMATION:
; APPLICANT: Wada, Masaru
; TITLE OF INVENTION: Levodione reductase gene and use thereof
; FILE REFERENCE: Levodione reductase
; CURRENT APPLICATION NUMBER: US/09/773,748
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: EP00101665.8
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; TYPE: PRT
; ORGANISM: Corynebacterium aquaticum
; US-09-773-748-1

Query Match 24.7%; Score 343.5; DB 9; Length 267;
Best Local Similarity 33.3%; Pred. No. 8.4e-25;
Matches 88; Conservative 48; Mismatches 105; Indels 23; Gaps 5;

QY 10 RLEGKVALITGGASGIGETAKLFSSQHGAKVAIADVQDE-----LGHSVVEAIGTSNSTYI 65
Db 10 RETDRVLLITGGSGGLGRATAVRLAAEGAKLSLDVDSSEGLEASKAAVLETPADAEVLTT 69
QY 66 HCDVTNEDGVKNVNDVTSTYKGLDIFMSNAGISDPNRPRIIDNEK---ADFERVLSVNV 122
Db 70 VADVSDEAQUEAVYVATTFERFRIDGFFNNAGIEGKNP-----TESFTAEEFDKVVYSINL 125
QY 123 TGVLFCMKHAARVMIPARSGNIISTASLSSTMTGGSSHAYCGSKHAVLALTRNLAVELGQ 182
Db 126 RGVFLGLEKVLKIMRQSGSMVNTASVGGIRGIGNQSGYAAAKHGCVGLTRNSAVEYGR 185
QY 183 FGIRVNCLSPFGLPTALGKKF-----SGIKNEEFENFVINFAGNLKGPKEFNVEDVANA 235
Db 186 YGIRINAIAPGAIWTPMVNSMKQLDPENPRKAAEEFIQV-----NPSKRYGEAPEIAAV 240
QY 236 ALYLASDEAKYVSGHNLFDGGFS 259
Db 241 VAPLLSDDASYNVATVVPIDGGQS 264

RESULT 6
US-09-940-037A-29
; Sequence 29, Application US/09940037A
; Patent No. US20020127679A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Kimoto, No. US20020127679A1Ihiro
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Mitsuhashi, Kazuya
; TITLE OF INVENTION: NOVEL CARBONYL REDUCTASE, METHOD FOR PRODUCING SAID ENZYME, DN
; TITLE OF INVENTION: ENCODING SAID ENZYME, AND METHOD FOR PRODUCING ALCOHOL USING
; FILE REFERENCE: ENZYME
; FILE REFERENCE: 06501-050001
; CURRENT APPLICATION NUMBER: US/09/940,037A
; CURRENT FILING DATE: 2000-08-27
; PRIOR APPLICATION NUMBER: 09/468,738
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: JP 1998-363130
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0, reformatted using WordPerfect 5.1
; SEQ ID NO 29
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-940-037A-29

Query Match 24.5%; Score 341.5; DB 10; Length 261;
Best Local Similarity 34.4%; Pred. No. 1.3e-24;
Matches 93; Conservative 50; Mismatches 98; Indels 29; Gaps 10;

QY 11 LEGKVALITGGASGIGETAKLFSSQHGAKVAI---ADVQD---ELGHSVVEAIGTSNSTYI 65
Db 5 LAGKVVAITGAASGLGKAMAIRFGKQAKVINYISNKQDPNEKVEKIKAGG--EAVVV 62
QY 66 HCDVTNEDGVKNVNDVTSTYKGLDIFMSNAGISDPNRPRIIDNEK--ADFERVLSVNV 123
Db 63 QGDVTKEEDVKNIVQTAIEFGFTLDIMINNAGLENP---VPSHEMPLKDWKVICGTNLT 118
QY 124 GYFLCMKHAARVMIPAR-SGNIISTASLSSTMTGGSSHAYCGSKHAVLALTRNLAVELGQ 182
Db 119 GAFGLSREAIKYFVENDIKGNINMSSVHEVWPWLFVHYAASKGKIKMTTLEALEYAP 178
QY 183 FGIRVNCLSPFGLPTAL-GKKFSGIKNEEFENFVINFAGNLKGPKEFNVEDVANAALYLAS 241
Db 179 KGI RVN NIGPGAINTPINAEKPADPKQKADVESMIPM-GYIGEP-----EIAA VAAWLAS 233
QY 242 DEAKYVSGHNLFDGGFSVCNSVIKVFQYP 271
Db 234 KEASYVTGITLFDGGMT-----QYP 254

RESULT 7
US-09-815-242-14092
; Sequence 14092, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14092
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-14092
```

```
Query Match 23.8%; Score 332; DB 10; Length 263;
Best Local Similarity 33.1%; Pred. No. 9.9e-24;
Matches 87; Conservative 49; Mismatches 113; Indels 14; Gaps 5;

QY 10 RLEGKVALITGASGIGETTAKLFSQHGAKVAIADVQDELGHVSVEAIGTSN-STYIHCD 68
Db 3 KLTGKTALITGASGIGEGIAFARHGANILLDISDEIEKLADELGGRGHRTAVKAD 62

QY 69 VNEDEGVKNVDNTVSTYTKLIDMFNSNAGISDPNRPRIIDNEKADFERVLSVNVTVGFLC 128
Db 63 VRDFASVQAAVARAKETEGRIDILVNAGVCRNLGNFLDMSEEDRDFH--IDINIKGVWNV 120

QY 129 MKHAARVMPARSGNIISTASLSSTM-GGSSSHAYCGSKHVALTRNLAVELGQFGIRV 187
Db 121 TRAVLPFEMIKRKGGRVMMSSVTGDMVADFGETAIALSKAAIVGLTKSLAVEYAQSGIRV 180

QY 188 NCLSPFGLPTALGKKFSGIKNEEFENVINPAG-----NLGPKFKNVEDVANAALYLAS 241
Db 181 NAICPGYVTRPMAESIAQSNPDPSVLTEMAKAIPRLRLADPL---EVGELAAFLAS 236

QY 242 DEAKVYSGHNLFDGGFVCSNV 264
Db 237 DESSYLTGTQNVIDGGSTLPESV 259
```

```
RESULT 8
US-09-815-242-13387
; Sequence 13387, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 13387
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13387

Query Match 23.8%; Score 331; DB 10; Length 271;
Best Local Similarity 33.6%; Pred. No. 1.3e-23;
Matches 85; Conservative 48; Mismatches 110; Indels 10; Gaps 6;

QY 11 LEGKVALITGASGIGETTAKLFSQHGAKVAIADVQDELGHVSVEA---ICTSNSTYIHC 67
Db 13 LKCKIALITGASYGIGFAIAKSYAAGATIVENDINQDLVKNKGEIAREVGEIANGIV-C 71

QY 68 DVNEDGVKNVDNTVSTYTKLIDMFNSNAGISDPNRPRIIDNEKADFERVLSVNVTVGFL 127
Db 72 DVTEDEGIQAMVKQISQEVGVVIDILVNAGII--RRVPMCEMSAADFKVIDIDLNAFFI 129

QY 128 CMKHAARVMPARSGNIISTASLSSTMGGSSSHAYCGSKHVALTRNLAVELGQFGIRV 187
Db 130 VSKAVIPSMIKKGHGCKIINICSMMSSELGRETVSAYAAKGGCKMLTRNIASEYGGANIQC 189

QY 188 NCLSP--FGLP-TALGKKFSGIKNEEFENVINPAGNLGPKFKNVEDVANAALYLASDEA 244
Db 190 NGIGPGYIATPQTAPLRELQEDGSRHPDFQFI-TAKTPAARWGNPDLMGPAVFLASDAS 248

QY 245 KYVSGHNLFDGG 257
Db 249 NFVNGHILYVDGG 261
```

```
RESULT 9
US-09-815-242-13679
; Sequence 13679, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13679
; LENGTH: 272
; TYPE: PRT
; ORGANISM: streptococcus pneumoniae
US-09-815-242-13679

Query Match 23.8%; Score 331; DB 10; Length 272;
Best Local Similarity 33.6%; Pred. No. 1.3e-23;
```

```
Db      69 FVGDVSSPSFVQATMKQLIVKNDGQIDILLNTAGTILDDYRPS - LETSALMDQIIATNLKS   127
Qy     125 VFLCMKHAAARVMIPARSNGNIISTASLSTMGSSHAYCGSKHVALTRNLAVELQGFG    184
        ||| : | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
Db     128 VFVTNALPYPFLOOKKGVIWNMASIAGLVAGGAAYTASKHAIGYTROLSVDYAKLG    187
Qy     185 IRVNCLSPFGLPTALGKKFKSGIKNEEFENVINPAGNK-----GPKNVEDVA    233
        || | |: | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     188 IRAWAIAPGAITPM-----NAADPAGEEMATWVASETPATGRWAQPQEVA    233
Qy     234 NAAILYLASDEAKYVSGHNFDFDGGFSV    260
        ::|||::|: | | | | | | | | | | | | | | | | | | | | | | |
Db     234 KLSLFSLADDADIHGTVMTIDGGWTM    260

RESULT 11
US-09-815-242-12688
; Sequence 12688, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlсен, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; ; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.Olla
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12688
; LENGTH: 272
; TYPE: PRt
; ORGANISM: Staphylococcus aureus
US-09-815-242-12688

Query Match          23.6%; Score 328.5; DB 10; Length 272;
Best Local Similarity 30.1%; Pred. No. 2,2e-23;
Matches              78; Conservative 63; Mismatches 99; Indels 19; Gaps 7;

Qy       9 RRLEKVALTGTGSAGTGTTAKLFSOHGAKVATAVDDELGHSV--VEAITGTSNSTYIH    66
         :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db      2 KRLENKVAVTVTGSTGTCQAIAIALOEGAVVLDAIEAVSETVDTIKSNGDKAKAVV-    60
Qy      67 CDVTNEDGVKNAVDTVSTYCKLDIMESNAGISDPNRPIIDNEKADEFVLSNVYTCVF    126
         ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:::::|:|:|:|
Db      61 VDIAEQOIDINFASEIREQFHGDVDFLNAGCV-DNAAGRHEYTPDYVDKI MNVDMRGTF    119
Qy     127 LCMKHAARVIP-----ARGSNIIISTASLSTMGSSHAYCGSKHAVLATRNLAVELCOF    183
         | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     120 LMWK----MWLPMLMTTKGSIIVNTSSFSGOADLYRSCYNNAKCAVINFTXSAIEYGRD    175
Qy     184 GIRVNCLSPFGLPTALGKKFKSGIKNEEFENVINPAGNLK-----GPKNVEDVANAAHY    238
         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```



```

; ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; ; Acids Encoding the Same
; ; FILE REFERENCE: P2630PIC11
; ; CURRENT APPLICATION NUMBER: US/09/978,295A
; ; CURRENT FILING DATE: 2001-10-15
; ; PRIOR APPLICATION NUMBER: 09/918585
; ; PRIOR FILING DATE: 2001-07-30
; ; PRIOR APPLICATION NUMBER: 60/062250

```

1	PRIOR APPLICATION NUMBER: 60/081955
2	PRIOR FILING DATE: 1998-04-15
3	PRIOR APPLICATION NUMBER: 60/081817
4	PRIOR FILING DATE: 1998-04-15
5	PRIOR APPLICATION NUMBER: 60/081819
6	PRIOR FILING DATE: 1998-04-15
7	PRIOR APPLICATION NUMBER: 60/081952
8	PRIOR FILING DATE: 1998-04-15
9	PRIOR APPLICATION NUMBER: 60/081838
10	PRIOR FILING DATE: 1998-04-15
11	PRIOR APPLICATION NUMBER: 60/082568
12	PRIOR FILING DATE: 1998-04-21
13	PRIOR APPLICATION NUMBER: 60/082569
14	PRIOR FILING DATE: 1998-04-21
15	PRIOR APPLICATION NUMBER: 60/082704
16	PRIOR FILING DATE: 1998-04-22
17	PRIOR APPLICATION NUMBER: 60/082804
18	PRIOR FILING DATE: 1998-04-22
19	PRIOR APPLICATION NUMBER: 60/082700
20	PRIOR FILING DATE: 1998-04-22
21	PRIOR APPLICATION NUMBER: 60/082797
22	PRIOR FILING DATE: 1998-04-22
23	PRIOR APPLICATION NUMBER: 60/082796
24	PRIOR FILING DATE: 1998-04-23
25	PRIOR APPLICATION NUMBER: 60/083336
26	PRIOR FILING DATE: 1998-04-27
27	PRIOR APPLICATION NUMBER: 60/083322
28	PRIOR FILING DATE: 1998-04-28
29	PRIOR APPLICATION NUMBER: 60/083392
30	PRIOR FILING DATE: 1998-04-29
31	PRIOR APPLICATION NUMBER: 60/083495
32	PRIOR FILING DATE: 1998-04-29
33	PRIOR APPLICATION NUMBER: 60/083496
34	PRIOR FILING DATE: 1998-04-29
35	PRIOR APPLICATION NUMBER: 60/083499
36	PRIOR FILING DATE: 1998-04-29
37	PRIOR APPLICATION NUMBER: 60/083545
38	PRIOR FILING DATE: 1998-04-29
39	PRIOR APPLICATION NUMBER: 60/083554
40	PRIOR FILING DATE: 1998-04-29
41	PRIOR APPLICATION NUMBER: 60/083558
42	PRIOR FILING DATE: 1998-04-29
43	PRIOR APPLICATION NUMBER: 60/083559
44	PRIOR FILING DATE: 1998-04-29
45	PRIOR APPLICATION NUMBER: 60/083500
46	PRIOR FILING DATE: 1998-04-29
47	PRIOR APPLICATION NUMBER: 60/083742
48	PRIOR FILING DATE: 1998-04-30
49	PRIOR APPLICATION NUMBER: 60/084366
50	PRIOR FILING DATE: 1998-05-05
51	PRIOR APPLICATION NUMBER: 60/084414
52	PRIOR FILING DATE: 1998-05-06
53	PRIOR APPLICATION NUMBER: 60/084441
54	PRIOR FILING DATE: 1998-05-06
55	PRIOR APPLICATION NUMBER: 60/084637
56	PRIOR FILING DATE: 1998-05-07
57	PRIOR APPLICATION NUMBER: 60/084639
58	PRIOR FILING DATE: 1998-05-07
59	PRIOR APPLICATION NUMBER: 60/084640
60	PRIOR FILING DATE: 1998-05-07
61	PRIOR APPLICATION NUMBER: 60/084627
62	PRIOR FILING DATE: 1998-05-07
63	PRIOR APPLICATION NUMBER: 60/084643
64	PRIOR FILING DATE: 1998-05-07
65	PRIOR APPLICATION NUMBER: 60/085339
66	PRIOR FILING DATE: 1998-05-13
67	PRIOR APPLICATION NUMBER: 60/085338
68	PRIOR FILING DATE: 1998-05-13
69	PRIOR APPLICATION NUMBER: 60/085323
70	PRIOR FILING DATE: 1998-05-13

[illegible]

Search completed: February 14, 2003, 09:25:30
Job time : 14 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 18:26:19 ; Search time 1468 seconds
(without alignments)
9035.493 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgttccaatccagattct 819

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_Other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	2.6	457	13	BM359844 GA_Ea002
2	21	2.6	598	12	BG442276 GA_Ea001
3	21	2.6	615	12	BF269379 GA_ER000
4	21	2.6	780	17	AZ211531 SP_0155_B
5	21	2.6	911	12	BG444414 GA_Ea002
6	20	2.4	474	17	AQ790030 HS_3193_B

7	20	2.4	597	14	BM690984
8	20	2.4	601	17	AQ735999
9	20	2.4	638	12	BG166718
10	20	2.4	670	13	BI669871
11	20	2.4	786	12	BG714518
12	20	2.4	894	10	BE571531
c 13	19	2.3	426	10	AW856509
c 14	19	2.3	436	10	AV969254
c 15	19	2.3	473	17	BH737850
c 16	19	2.3	485	17	AQ185426
c 17	19	2.3	506	17	AQ514891
c 18	19	2.3	509	10	AW922556
c 19	19	2.3	515	12	BG157985
c 20	19	2.3	528	17	BH451022
c 21	19	2.3	537	13	BM318747
c 22	19	2.3	541	17	AZ622455
c 23	19	2.3	584	17	BH494130
c 24	19	2.3	587	17	BH555089
c 25	19	2.3	592	14	BP016339
c 26	19	2.3	593	12	BG048039
c 27	19	2.3	600	13	BM076927
c 28	19	2.3	600	17	AZ591300
c 29	19	2.3	635	17	AZ878058
c 30	19	2.3	643	17	AZ389227
c 31	19	2.3	645	17	BH569302
c 32	19	2.3	663	17	AQ315681
c 33	19	2.3	675	17	BH569310
c 34	19	2.3	700	14	BQ111969
c 35	19	2.3	702	14	BQ873587
c 36	19	2.3	704	14	BQ875148
c 37	19	2.3	705	17	BH603059
c 38	19	2.3	716	14	BQ972114
c 39	19	2.3	717	14	BQ971515
c 40	19	2.3	738	14	BQ870873
c 41	19	2.3	788	14	BQ506405
c 42	19	2.3	800	17	CNS034HP
c 43	19	2.3	809	17	AG160498
c 44	19	2.3	811	17	CNS0482C
c 45	19	2.3	821	17	CNS02LJ9

ALIGNMENTS

RESULT 1

BM359844

LOCUS

DEFINITION

GA_Ea0024E17r Gossypium arboreum 7-10 dpa fiber library Gossypium

ARBOREUM cDNA clone GA_Ea0024E17r, mRNA sequence.

ACCESSION

BM359844

VERSION

BM359844.1

KEYWORDS

EST

SOURCE

Gossypium arboreum

ORGANISM

Gossypium arboreum

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM359844 457 bp mRNA linear EST 09-JAN-2002
GA_Ea0024E17r Gossypium arboreum 7-10 dpa fiber library Gossypium
ARBOREUM cDNA clone GA_Ea0024E17r, mRNA sequence.

BM359844.1 GI:18100590

EST

Gossypium arboreum

Gossypium arboreum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

1 (bases 1 to 457)

Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry

, D., Wood, T.C., Leslie, A. and Wilkins, T.A.

An integrated analysis of the genetics, development, and evolution

of the cotton fiber

Unpublished (2000)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total High Quality bases = 418

Seq primer: TAATACGACTCACTAGG

High quality sequence start: 5

FEATURES
source High quality sequence stop: 451.
Location/Qualifiers
1. .457
/organism="Gossypium arboreum"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0024E17r"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 116 a 131 c 122 g 87 t 1 others
ORIGIN
Query Match 2.6%; Score 21; DB 13; Length 457;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CTACATCCACTGTGATGTTAC 209
|||||
Db 169 CTACATCCACTGTGATGTTAC 189

RESULT 2
BG442276 598 bp mRNA linear EST 15-MAR-2001
LOCUS GA_Ea0016G23f Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION arboreum cDNA clone GA_Ea0016G23f, mRNA sequence.
ACCESSION BG442276
VERSION BG442276.1 GI:13351928
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 598)
AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGG
High quality sequence stop: 432.

FEATURES
source 1. .598
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0016G23f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 161 a 119 c 200 g 117 t 1 others
ORIGIN
Query Match 2.6%; Score 21; DB 12; Length 598;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CATTGCTGATGTCGAAGATGA 143

Db 316 CATTGCTGATGTCGAAGATGA 336
|||||

RESULT 3
BF269379 615 bp mRNA linear EST 07-MAR-2001
LOCUS GA_EB0004I11f Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION arboreum cDNA clone GA_EB0004I11f, mRNA sequence.
ACCESSION BF269379
VERSION BF269379.1 GI:11200374
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 615)
AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGG
High quality sequence stop: 568.

FEATURES
source 1. .615
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_EB0004I11f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 151 a 166 c 176 g 120 t 2 others
ORIGIN
Query Match 2.6%; Score 21; DB 12; Length 615;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CTACATCCACTGTGATGTTAC 209
|||||

Db 184 CTACATCCACTGTGATGTTAC 204

RESULT 4
AZ211531 780 bp DNA linear GSS 31-AUG-2000
LOCUS SF_0155_B1_C09_SP6E Strongylocentrotus purpuratus, purple sea
DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus
genomic clone Plate=155 Col=17 Row=F, DNA sequence.
ACCESSION AZ211531
VERSION AZ211531.1 GI:8427406
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus

REFERENCE Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 780)
AUTHORS Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray

TITLE
JOURNAL
MEDLINE
COMMENT

G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17): 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 155 row: F column: 17
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 780.

FEATURES
source
Location/Qualifiers
1. .780
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate:155 Col=17 Row=F"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 326 a 126 c 151 g 177 t
ORIGIN

Query Match 2.6%; Score 21; DB 17; Length 780;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 607 TTTTCAGGGATTAAATGAA 627
|||||
Db 97 TTTTCAGGGATTAAATGAA 117

RESULT 5
BG444414 911 bp mRNA linear EST 15-MAR-2001
LOCUS
DEFINITION
GA_Ea0024E17f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea0024E17f, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

EST.
GI:13354066
Gossypium arboreum.
Gossypium arboreum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 911)
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry D., Wood, T.C., Leslie, A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution of the cotton fiber
Unpublished (2000)
Contact: Wing RA
Clemson University
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCATATAGG
High quality sequence stop: 650.

FEATURES
source
Location/Qualifiers
1. .911
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0024E17f"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 233 a 207 c 280 g 190 t
ORIGIN

Query Match 2.6%; Score 21; DB 12; Length 911;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CTACATCCACTGTGATGTTAC 209
|||||
Db 170 CTACATCCACTGTGATGTTAC 190

RESULT 6
AQ790030 474 bp DNA linear GSS 03-AUG-1999
LOCUS
DEFINITION
HS_3193_B1_E04_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3193 Col=7 Row=J, DNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AQ790030
AQ790030.1 GI:5697654
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 474)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3193 row: J column: 7
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 474.

FEATURES
source
Location/Qualifiers
1. 474
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3193 Col=7 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 135 a 78 c 76 g 182 t
ORIGIN

Query Match 2.4%; Score 20; DB 17; Length 474;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 TCAACTATGGTGGTGGTTC 473
|||||
Db 300 TCAACTATGGTGGTGGTTC 319

RESULT 7
BM690984


```
Plate: LLAM10227 row: d column: 10
High quality sequence stop: 636.
FEATURES
source
1. .638
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_89"
/tissue_type="hyponephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 215 a 105 c 111 g 207 t
ORIGIN
Query Match 2.4%; Score 20; DB 12; Length 638;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 TCAACCTATGGAACCTGGA 269
|||||
Db 469 TCAACCTATGGAACCTGGA 488
|||||
RESULT 10
LOCUS BI669871 670 bp mRNA linear EST 12-SEP-2001
DEFINITION 603293467F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5312657 5',
mRNA sequence.
ACCESSION BI669871
VERSION BI669871.1 GI:15584104
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 670)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11792 row: f column: 18
High quality sequence stop: 670.
FEATURES
source
1. .670
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 213 a 123 c 126 g 208 t
ORIGIN
Query Match 2.4%; Score 20; DB 12; Length 786;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 TCAACCTATGGAACCTGGA 269
|||||
Db 431 TCAACCTATGGAACCTGGA 450
|||||
RESULT 12
LOCUS BE571531 894 bp mRNA linear EST 15-AUG-2000
DEFINITION 601333976F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3711403 5',
mRNA sequence.
```

```

ACCESSION   BE571531
VERSION     BE571531.1  GI:9815251
SOURCE      house mouse.
ORGANISM    Mus musculus

REFERENCE   1 (bases 1 to 894)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Jeffrey Green M.D.
           cDNA Library Preparation: Life Technologies, Inc.
           CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: L1AM8935 row: 9 column: 20
           High quality sequence stop: 651.
           Location/Qualifiers
             1..894
               /organism="Mus musculus"
               /strain="FVB/N"
               /db_xref="taxon:10090"
               /clone="IMAGE:3711403"
               /clone_lib="NCI_CGAP_Mam6"
               /sex="female, virgin"
               /tissue_type="infiltrating ductal carcinoma"
               /dev_stage="5 months"
               /lab_host="DH10B"
               /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
               Site_2: NotI; Cloned unidirectionally. Primer: Oligo dn.
               Library constructed by Life Technologies. Investigator
               providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT  278 a 127 c 190 g 299 t
ORIGIN
Query Match      . 2.4%; Score 20; DB 10; Length 894;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  630 AGAATTTGAGATGTAATAA 649
Db  242 AGAATTTGAGATGTAATAA 261
|||||
|||||

RESULT 13
ACCESSION   AW856509
LOCUS       AW856509/c 346 bp mRNA linear EST 19-MAY-2000
DEFINITION  PM4-CT0291-271099-001-d02 CT0291 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW856509
VERSION     AW856509.1  GI:7952202
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 346)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics

```

```

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-pM4-CT0291-271
099-001-d02&t3=1999-10-27&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 25
High quality sequence stop: 137.
Location/Qualifiers
  1..346
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="CT0291"
    /dev_stage="Adult"
    /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
    SmaI; A mini-library was made by cloning products derived
    from ORESTES PCR (U.S. Letters Patent application No. 196
    ,716 - Ludwig Institute for Cancer Research) profiles
    into the puc 18 vector. Reverse transcription of tissue
    mRNA and cDNA amplification were performed under low
    stringency conditions."

BASE COUNT  65 a 82 c 96 g 103 t
ORIGIN
Query Match      2.3%; Score 19; DB 10; Length 346;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  109 GGAGCCAAAGTTGCCATTG 127
Db  113 GGAGCCAAAGTTGCCATTG 95
|||||
|||||

RESULT 14
ACCESSION   AV969254
LOCUS       AV969254 426 bp mRNA linear EST 14-MAR-2002
DEFINITION  AV969254 Nori Satoh unpublished cDNA library, larva Ciona
intestinalis cDNA clone cily18g07 5', mRNA sequence.
ACCESSION   AV969254
VERSION     AV969254.1  GI:19459018
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
ORGANISM    Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Clona.
1 (bases 1 to 426)
Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoheascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
  1..426
    /organism="Ciona intestinalis"
    /db_xref="taxon:7719"
    /clone="cily18g07"
    /clone_lib="Nori Satoh unpublished cDNA library, larva"
    /tissue_type="whole animal"
    /dev_stage="larva"
    /note="Vector: pBluescript SK"

BASE COUNT  148 a 81 c 91 g 106 t
ORIGIN

```

Query Match 2.3%; Score 19; DB 10; Length 426;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 648 AACTTTTCGGGAATTG 666
|||||
Db 367 AACTTTTCGGGAATTG 385

RESULT 15
BH737850
LOCUS BH737850 473 bp DNA linear GSS 20-FEB-2002
DEFINITION BOHZN64TR BO_2_3_KB Brassica oleracea genomic clone BOHZN64, DNA
sequence.
ACCESSION BH737850
VERSION BH737850.1 GI:18843245
KEYWORDS GSS
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 473)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOHZN64TF
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: Sheared ends.
FEATURES
Location/Qualifiers
1..473
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOHZN64"
/clone_lib="BO_2_3_KB"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"

BASE COUNT 147 a 76 c 97 g 153 t
ORIGIN

Query Match 2.3%; Score 19; DB 17; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 AGAAATTTTCAGGATTAA 620
|||||
Db 192 AGAAATTTTCAGGATTAA 210

Search completed: February 18, 2003, 19:41:21
Job time : 1477 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 18:30:54 ; Search time 80 seconds
(without alignments)
3139.603 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgtccaataatccagattctt 819

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	2.2	511	US-09-404-879A-64	Sequence 64, Appl
2	17	2.1	842	US-08-998-416-312	Sequence 312, App
3	17	2.1	3388	US-09-141-206-1	Sequence 1, Appli
4	17	2.1	3456	US-09-134-001C-405	Sequence 405, App
5	17	2.1	20137	US-09-262-773-206	Sequence 206, App
6	17	2.1	20138	US-09-262-773-9	Sequence 9, Appli
7	17	2.1	23071	US-09-262-773-210	Sequence 210, App
8	16	2.0	426	US-08-470-179-116	Sequence 116, App
9	16	2.0	480	US-09-134-001C-845	Sequence 845, App
10	16	2.0	515	US-09-222-575-143	Sequence 143, App
11	16	2.0	615	US-09-328-111-197	Sequence 197, App
12	16	2.0	861	US-09-134-001C-1266	Sequence 1266, Ap
13	16	2.0	1019	US-08-225-473-2	Sequence 2, Appli
14	16	2.0	1423	US-08-948-616-4	Sequence 4, Appli
15	16	2.0	1423	US-09-193-510-4	Sequence 4, Appli
16	16	2.0	1423	US-09-368-402-4	Sequence 4, Appli
17	16	2.0	1425	US-08-883-515-1	GENERAL INFORMAT
18	16	2.0	1839	US-09-134-001C-1191	Sequence 1191, Ap
19	16	2.0	2236	US-08-484-993B-13	Sequence 13, Appl
20	16	2.0	2236	US-08-484-158B-13	Sequence 13, Appl
21	16	2.0	2236	US-08-484-596A-13	Sequence 13, Appl
22	16	2.0	2236	US-08-480-150A-13	Sequence 13, Appl
23	16	2.0	2236	US-08-458-731-13	Sequence 13, Appl
24	16	2.0	2236	US-08-149-223A-13	Sequence 13, Appl
25	16	2.0	2237	US-09-156-253-1	Sequence 1, Appli
26	16	2.0	2558	US-08-999-733-3	Sequence 3, Appli
27	16	2.0	2571	US-07-718-575-7	Sequence 7, Appli

28	16	2.0	2971	1	US-08-481-206-7	Sequence 7, Appli
29	16	2.0	2971	2	US-08-486-269A-7	Sequence 7, Appli
30	16	2.0	3386	1	US-08-703-809-2	Sequence 2, Appli
31	16	2.0	3386	2	US-08-703-808-2	Sequence 2, Appli
32	16	2.0	3386	1	US-08-914-068-2	Sequence 2, Appli
33	16	2.0	3386	2	US-08-703-807-2	Sequence 2, Appli
34	16	2.0	3386	2	US-08-747-108A-2	Sequence 2, Appli
35	16	2.0	3386	3	US-09-211-631-2	Sequence 2, Appli
36	16	2.0	3386	3	US-09-265-628-2	Sequence 2, Appli
37	16	2.0	3386	4	US-09-001-141-2	Sequence 2, Appli
38	16	2.0	3386	4	US-09-532-803-2	Sequence 2, Appli
39	16	2.0	9103	1	US-08-466-033-182	Sequence 182, App
40	16	2.0	9103	1	US-08-444-733-182	Sequence 182, App
41	16	2.0	9103	2	US-08-464-134-182	Sequence 182, App
42	16	2.0	9103	2	US-08-461-361-182	Sequence 182, App
43	16	2.0	9103	2	US-08-485-910-182	Sequence 182, App
44	16	2.0	9103	5	PCT-US95-06286-156	Sequence 156, App
45	16	2.0	9757	1	US-08-093-453B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-404-879A-64
; Sequence 64, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-64

Query Match 2.2%, Score 18; DB 4; Length 511;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAGGAGGAGCCAGTGGAA 73
|||||
DB 389 CAGGAGGAGCCAGTGGAA 406

RESULT 2
US-08-998-416-312
; Sequence 312, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina

; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 20138
; TYPE: DNA
; ORGANISM: human
US-09-262-773-9

Query Match 2.1%; Score 17; DB 4; Length 20138;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 GTGTTCCATATCCAGA 815
|||||
Db 10850 GTGTTCCATATCCAGA 10834

RESULT 7
US-09-262-773-210/c
; Sequence 210, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 210
; LENGTH: 23071
; TYPE: DNA
; ORGANISM: human
US-09-262-773-210

Query Match 2.1%; Score 17; DB 4; Length 23071;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 GTGTTCCATATCCAGA 815
|||||
Db 13783 GTGTTCCATATCCAGA 13767

RESULT 8
US-08-470-179-116/c
; Sequence 116, Application US/08470179
; Patent No. 5645994
; GENERAL INFORMATION:
; APPLICANT: Huang Ph.D, Wai Mun
; TITLE OF INVENTION: Method and Compositions for
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Trask, Britt and Rossa
; STREET: P.O. Box 2550
; CITY: Salt Lake City
; STATE: Utah
; COUNTRY: USA
; ZIP: 84110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,179
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sweigert Ph.D, Susan E.
; REGISTRATION NUMBER: 36,289
; REFERENCE/DOCKET NUMBER: 2601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 801-532-1922
; TELEFAX: 801-531-9168
; INFORMATION FOR SEQ ID NO: 116:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 426 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Anacystis nidulans
US-08-470-179-116

Query Match 2.0%; Score 16; DB 1; Length 426;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GAGCCAAAGTTGCCAT 125
|||||
Db 218 GAGCCAAAGTTGCCAT 203

RESULT 9
US-09-134-001C-845
; Sequence 845, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 845
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-845

Query Match 2.0%; Score 16; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 ATCAAAAGTTGCCAT 808
|||||
Db 335 ATCAAAAGTTGCCAT 350

RESULT 10
US-09-222-575-143/c
; Sequence 143, Application US/09222575
; Patent No. 6387697
; GENERAL INFORMATION:
; APPLICANT: Yuqiu, Jiang
; APPLICANT: Dillon, Davin C.

```
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 210121.470
; CURRENT APPLICATION NUMBER: US/09/222,575
; CURRENT FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 143
; LENGTH: 515
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (5)
; OTHER INFORMATION: where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (286)
; OTHER INFORMATION: where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (498)
; OTHER INFORMATION: where n is a, c, g or t
; US-09-222-575-143

Query Match          2.0%; Score 16; DB 4; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 ATGAAGAAGAAATTGCA 638
Db 281 ATGAAGAAGAAATTGCA 266
|||||

RESULT 11
US-09-328-111-197/c
; Sequence 197, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Cartoll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 197
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(615)
; OTHER INFORMATION: n = A,T,C or G
; US-09-328-111-197

Query Match          2.0%; Score 16; DB 4; Length 615;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 271 ATTATGTTCCAGCAATG 286
Db 261 ATTATGTTCCAGCAATG 246
|||||

RESULT 12
US-09-134-001C-1266
; Sequence 1266, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1266
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-1266

Query Match          2.0%; Score 16; DB 4; Length 861;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 ATTAATAATGAAGAAG 631
Db 676 ATTAATAATGAAGAAG 691
|||||

RESULT 13
US-08-225-473-2/c
; Sequence 2, Application US/08225473
; Patent No. 5610012
; GENERAL INFORMATION:
; APPLICANT: Luchansky, John B.
; APPLICANT: Chen, Jianchi
; TITLE OF INVENTION: DNA Probes Specific For Virulent
; TITLE OF INVENTION: Listeria Monocytogenes
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,473
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muetting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 800.16-US-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1019 base pairs
```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-225-473-2

Query Match 2.0%; Score 16; DB 1; Length 1019;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AACTGCATTCGCAAGA 27
|||||
Db 157 AACTGCATTCGCAAGA 142

RESULT 14

US-08-948-616-4
; Sequence 4, Application US/08948616
; Patent No. 5840539

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Lal, Preeti

; APPLICANT: Shah, Purvi

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/948,616

; FILING DATE: Herewith

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0409 US

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1423 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: LUNGUT07

; CLONE: 2607662

US-08-948-616-4

Query Match 2.0%; Score 16; DB 2; Length 1423;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 GAAACTGGACATTAT 275

|||||

Db 204 GAAACTGGACATTAT 219

RESULT 15

US-09-193-510-4

; Sequence 4, Application US/09193510

; Patent No. 5981226

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Lal, Preeti

; APPLICANT: Shah, Purvi

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/193,510

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/948,616

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0409 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1423 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: LUNGUT07

; CLONE: 2607662

US-09-193-510-4

Query Match 2.0%; Score 16; DB 2; Length 1423;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 GAAACTGGACATTAT 275

|||||

Db 204 GAAACTGGACATTAT 219

Search completed: February 18, 2003, 19:43:18
Job time : 113 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 19:33:55 ; Search time 114 Seconds
(without alignments)
3659.088 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgtccaatatccagattct 819

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 424239 seqs, 254661826 residues

Word size : 0

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	11.1	4505	9 US-09-944-160-23	Sequence 23, Appl
2	21	2.6	21	9 US-09-944-160-33	Sequence 33, Appl
3	21	2.6	21	9 US-09-944-160-43	Sequence 43, Appl
4	19	2.3	386	10 US-09-960-352-14577	Sequence 14577, A
5	18	2.2	292	9 US-09-796-692-5749	Sequence 5749, Ap
6	18	2.2	480	10 US-09-867-701-1789	Sequence 1789, Ap
7	18	2.2	511	10 US-09-884-441-64	Sequence 64, Appl
8	18	2.2	2016	10 US-09-764-870-237	Sequence 237, App
9	18	2.2	2223	10 US-09-764-870-110	Sequence 110, App
10	18	2.2	10785	10 US-09-764-878-324	Sequence 324, App
11	18	2.2	13315	10 US-09-764-878-328	Sequence 328, App
12	18	2.2	20261	10 US-09-764-878-325	Sequence 325, App
13	18	2.2	35641	10 US-09-962-436-306	Sequence 306, App
14	18	2.2	35641	10 US-09-880-107-2225	Sequence 2225, Ap
15	17	2.1	240	10 US-09-923-876-3491	Sequence 3491, Ap
16	17	2.1	356	10 US-09-960-352-12402	Sequence 12402, A
17	17	2.1	425	10 US-09-880-107-2010	Sequence 2010, Ap
18	17	2.1	517	10 US-09-880-107-1511	Sequence 1511, Ap
19	17	2.1	554	10 US-09-864-761-11940	Sequence 11940, A

c 20	17	2.1	596	10 US-09-864-761-16202	Sequence 16202, A
c 21	17	2.1	708	10 US-09-770-149-223	Sequence 223, App
c 22	17	2.1	756	9 US-09-774-639-64	Sequence 64, Appl
c 23	17	2.1	1438	10 US-09-884-901-1	Sequence 1, Appl
c 24	17	2.1	5313	10 US-09-801-368-297	Sequence 297, App
c 25	17	2.1	16423	9 US-09-989-442-160	Sequence 160, App
c 26	17	2.1	38059	10 US-09-880-107-2125	Sequence 2125, Ap
c 27	17	2.1	513509	9 US-09-754-853A-4	Sequence 4, Appl
c 28	16	2.0	100	10 US-09-864-761-32694	Sequence 32694, A
c 29	16	2.0	213	10 US-09-878-574-5029	Sequence 5029, Ap
c 30	16	2.0	270	10 US-09-878-574-12477	Sequence 12477, A
c 31	16	2.0	271	10 US-09-878-574-5697	Sequence 5697, App
c 32	16	2.0	291	10 US-09-878-574-7983	Sequence 7983, Ap
c 33	16	2.0	300	10 US-09-783-590-1117	Sequence 1117, Ap
c 34	16	2.0	319	10 US-09-864-761-23571	Sequence 23571, A
c 35	16	2.0	328	10 US-09-969-708-572	Sequence 572, App
c 36	16	2.0	373	10 US-09-983-965-4591	Sequence 4591, Ap
c 37	16	2.0	379	10 US-09-925-239-124	Sequence 124, App
c 38	16	2.0	380	10 US-09-983-965-215	Sequence 215, App
c 39	16	2.0	401	9 US-10-025-380-113	Sequence 113, App
c 40	16	2.0	401	10 US-09-922-217-113	Sequence 113, App
c 41	16	2.0	401	10 US-09-833-263-113	Sequence 113, App
c 42	16	2.0	401	10 US-09-878-574-3924	Sequence 3924, Ap
c 43	16	2.0	411	10 US-09-878-574-4099	Sequence 4099, Ap
c 44	16	2.0	412	9 US-09-954-531-790	Sequence 790, App
c 45	16	2.0	412	10 US-09-954-456-893	Sequence 893, App

ALIGNMENTS

RESULT 1
US-09-944-160-23
; Sequence 23, Application US/09944160
; Patent No. US20020174452A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, No. US20020174452A1man
; APPLICANT: Davin, Laurence
; APPLICANT: .. Huang, Ning
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan
; TITLE OF INVENTION: Content
; FILE REFERENCE: WSUR117983
; CURRENT APPLICATION NUMBER: US/09/944,160
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 60/230,632
; PRIOR FILING DATE: 2000-09-07
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 4505
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid pAPI249 encoding secoisolariciresinol
; OTHER INFORMATION: dehydrogenase (dehydrogenase)
US-09-944-160-23

Query Match 11.1%; Score 91; DB 9; Length 4505;
Best Local Similarity 100.0%; Pred.No. 9.2e-40;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CACAGCAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGCATGCCAAGATGA 143

Db 1045 CACAGCAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGCATGCCAAGATGA 1104

QY 144 ATTAGTCACTCAGTTGTTCGAGGCCATTGGC 174

Db 1105 ATTAGTCACTCAGTTGTTCGAGGCCATTGGC 1135

RESULT 2

US-09-944-160-33/c

; Sequence 33, Application US/09944160

```
; Patent No. US20020174452A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, No. US20020174452Alman
; APPLICANT: David, Laurence
; APPLICANT: Huang, Ning
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan
; FILE REFERENCE: WSURL17983
; CURRENT APPLICATION NUMBER: US/09/944,160
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 60/230,632
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
; US-09-944-160-33

Query Match      2.6%; Score 21; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GGGTGGTGGTCTTCACATGC 482
DB 21 GGGTGGTGGTCTTCACATGC 1

RESULT 3
US-09-944-160-43/c
; Sequence 43, Application US/09944160
; Patent No. US20020174452A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, No. US20020174452Alman
; APPLICANT: David, Laurence
; APPLICANT: .. Huang, Ning
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan
; FILE REFERENCE: WSURL17983
; CURRENT APPLICATION NUMBER: US/09/944,160
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 60/230,632
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
; US-09-944-160-43

Query Match      2.6%; Score 21; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GGGTGGTGGTCTTCACATGC 482
DB 21 GGGTGGTGGTCTTCACATGC 1

RESULT 4
US-09-960-352-14577/c
; Sequence 14577, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Rao, Nengbing
; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14577
; LENGTH: 386
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 62-LIB34-012-Q1-E1-H10
; US-09-960-352-14577

Query Match      2.3%; Score 19; DB 10; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 ATCAAGTGTTCCAATATC 811
DB 46 ATCAAGTGTTCCAATATC 28

RESULT 5
US-09-796-692-5749/c
; Sequence 5749, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5749
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-796-692-5749

Query Match      2.2%; Score 18; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 AAGCTAGAGGAAAGT 44
```



```
Db 42 AAGGCTAGAGGAAAGT 25
|||||
RESULT 6
US-09-867-701-1789
; Sequence 1789, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; OTHER INFORMATION: AND DIAGNOSIS OF OVARIAN CANCER
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1789
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(480)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-1789

Query Match 2.2%; Score 18; DB 10; Length 480;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGAATT 75
|||||
Db 424 GGAGGAGCCAGTGAATT 441

RESULT 7
US-09-884-441-64
; Sequence 64, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C7
; OTHER INFORMATION: DIAGNOSIS OF OVARIAN CANCER
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-884-441-64

Query Match 2.2%; Score 18; DB 10; Length 511;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAGGAGGAGCCAGTGGAA 73
|||||
Db 389 CAGGAGGAGCCAGTGGAA 406

RESULT 8
US-09-764-870-237/c
; Sequence 237, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
US-09-764-870-237/c

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 237
; LENGTH: 2016
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (93)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-870-237

Query Match 2.2%; Score 18; DB 10; Length 2016;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGAATT 75
|||||
Db 1098 GGAGGAGCCAGTGAATT 1081

RESULT 9
US-09-764-870-110/c
; Sequence 110, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 110
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-870-110/c

Query Match 2.2%; Score 18; DB 10; Length 2223;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGAATT 75
|||||
Db 1249 GGAGGAGCCAGTGAATT 1232

RESULT 10
US-09-764-878-324/c
; Sequence 324, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
US-09-764-878-324/c
```

```
; SEQ ID NO 324
; LENGTH: 10785
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-324

Query Match      2.2%; Score 18; DB 10; Length 10785;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAATT 75
Db 9833 GGAGGAGCCAGTGGAAATT 9816
|||||

RESULT 11
US-09-764-878-328/c
; Sequence 328, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 328
; LENGTH: 13315
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-328

Query Match      2.2%; Score 18; DB 10; Length 13315;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAATT 75
Db 12363 GGAGGAGCCAGTGGAAATT 12346
|||||

RESULT 12
US-09-764-878-325/c
; Sequence 325, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 325
; LENGTH: 20261
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-325

Query Match      2.2%; Score 18; DB 10; Length 20261;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAATT 75
Db 19310 GGAGGAGCCAGTGGAAATT 12993
|||||

RESULT 13
US-09-962-436-306

; Sequence 306, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 306
; LENGTH: 35641
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-306

Query Match      2.2%; Score 18; DB 10; Length 35641;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 608 TTTCAGGAGTAAAAATG 625
Db 19637 TTTCAGGAGTAAAAATG 19654
|||||

RESULT 14
US-09-880-107-2225
; Sequence 2225, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2225
; LENGTH: 35641
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L76568
US-09-880-107-2225

Query Match      2.2%; Score 18; DB 10; Length 35641;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 608 TTTCAGGAGTAAAAATG 625
Db 19637 TTTCAGGAGTAAAAATG 19654
|||||

RESULT 15
US-09-923-876-3491/c
; Sequence 3491, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
```

APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
FILE REFERENCE: PL-0012-1 CON
CURRENT APPLICATION NUMBER: US/09/923,876
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/298,329
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/085,331
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 6332
SOFTWARE: PERL Program
SEQ ID NO 3491
LENGTH: 240
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20020013958A1 700162378H1
US-09-923-876-3491

Query Match 2.1%; Score 17; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 TTCAGCAATGCAGGAAT 293
|||||
Db 92 TTCAGCAATGCAGGAAT 76

Search completed: February 18, 2003, 20:29:04
Job time : 160 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 17:45:44 ; Search time 250 Seconds
(without alignments)
7377.546 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgatt.....tggccaatatccagattctt 819

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	819	100.0	819	21	AAZ45730
2	816	99.6	816	21	AAZ45743
3	647	79.0	831	21	AAZ45731
4	95	11.6	819	21	AAZ45732
5	91	11.1	828	21	AAZ45734
6	91	11.1	4505	24	AAZ38432
7	87	10.6	831	21	AAZ45733
8	29	3.5	33	21	AAZ45744
9	23	2.8	23	21	AAZ45741

c	10	21	24	2.6	21	24	AAZ38441	Dehydrogenase dete
c	11	21	24	2.6	21	24	AAZ38451	Gt1 dehydrogenase
c	12	19	21	2.3	703	21	AAZ12470	Aspergillus oryzae
c	13	19	24	2.3	1029	24	ABQ68055	Listeria monocytog
c	14	19	24	2.3	1029	24	ABQ68055	Listeria monocytog
c	15	18	24	2.2	480	24	ABL78811	Human ovarian can
c	16	18	24	2.2	493	21	AAA43922	Human secreted exp
c	17	18	24	2.2	503	24	ABL62389	Colon adenocarcino
c	18	18	24	2.2	511	21	AAA69754	Human ovarian carc
c	19	18	24	2.2	511	24	ABN72648	Ovarian carcinoma
c	20	18	24	2.2	519	24	ABN65725	Human cancer relat
c	21	18	24	2.2	538	22	AAK92579	Human CDNA 3'-end
c	22	18	24	2.2	547	24	ABN63870	Human cancer relat
c	23	18	24	2.2	569	20	AAK97648	Extended human sec
c	24	18	24	2.2	577	23	ABV08508	Human prostate exp
c	25	18	24	2.2	1120	22	AAH94561	Human foetal CDNA,
c	26	18	24	2.2	1140	24	AAK97159	Human aspartyl pro
c	27	18	24	2.2	1361	24	AAK38827	Human PSNA CDNA #7
c	28	18	24	2.2	1364	22	AAK40957	CDNA encoding nove
c	29	18	24	2.2	1454	23	ABL29517	Drosophila melanog
c	30	18	24	2.2	1839	22	ABA08975	Human granuphilin-
c	31	18	24	2.2	1909	22	AAK94727	Human full-length
c	32	18	24	2.2	2016	22	AAK31423	Human CDNA encodin
c	33	18	24	2.2	2016	23	ABK43880	DNA encoding novel
c	34	18	24	2.2	2016	24	ABQ66747	Human polynucleoti
c	35	18	24	2.2	2223	22	AAK31296	Human CDNA encodin
c	36	18	24	2.2	2223	24	ABQ66620	Human polynucleoti
c	37	18	24	2.2	2356	22	AAK09392	Human vesicle traf
c	38	18	24	2.2	2704	24	ABK88007	CDNA encoding huma
c	39	18	24	2.2	3767	22	AAK63413	Human immune/haema
c	40	18	24	2.2	3865	22	AAK63414	Human immune/haema
c	41	18	24	2.2	3878	23	ABL29516	Drosophila melanog
c	42	18	24	2.2	4007	24	ABQ54678	Human ovarian anti
c	43	18	24	2.2	4156	23	ABL29526	Drosophila melanog
c	44	18	24	2.2	4285	23	ABL25336	Drosophila melanog
c	45	18	24	2.2	4651	24	ABK93534	Human breast speci

ALIGNMENTS

RESULT 1	AAZ45730	standard; CDNA; 819 BP.
ID	AAZ45730	standard; CDNA; 819 BP.
AC	AAZ45730;	
XX	06-APR-2000	(first entry)
DT	Secoisolaricresinol dehydrogenase CDNA clone DERV133.	
DE	Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;	
XX	lignan; matairesinol; health-protecting lignan; phytoestrogen;	
KW	enterolactone; enterodiol; neutriceutical; dietary supplement;	
KW	(-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.	
XX	Forsythia intermedia.	
OS		
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..819
FT		/tag= a
FT		/product= "secoisolaricresinol dehydrogenase"
FT		/note= "no termination codon given"
XX		
PN	WO955846-A1.	
XX		
PD	04-NOV-1999.	
XX		
PF	23-APR-1999;	99WO-US08975.
XX		
PR	24-APR-1998;	98US-0082977.
XX		
PA	(UNIW) UNIV WASHINGTON STATE RES' FOUND.	

Db 264 ACTGGACATTATGTTCA 280

RESULT 5

AAZ45734

ID AAZ45734 standard; cDNA; 828 BP.

XX

AC AAZ45734;

XX

DT 06-APR-2000 (first entry)

XX

DE Secoisolariciresinol dehydrogenase cDNA clone SMDEHY631.

XX

KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway; lignan; matairesinol; health-protecting lignan; phytoestrogen; enterolactone; enterodiol; neutriceutical; dietary supplement; (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.

XX

OS Forsythia intermedia.

XX

FH Key Location/Qualifiers

FT CDS 1..828

FT /*tag= a

FT /product= "secoisolariciresinol dehydrogenase"

FT /note= "no termination codon given"

XX

PN WO9955846-A1.

XX

PD 04-NOV-1999.

XX

PF 23-APR-1999; 99WO-US089975.

XX

PR 24-APR-1998; 98US-0082977.

XX

PA (UNIW) UNIV WASHINGTON STATE RES FOUND.

XX

PI Xia Z, Costa MA, Davin LB, Lewis NG;

XX

DR WPI; 2000-126356/11.

XX

DR P-PSDB; AAY54416.

XX

PT New nucleic acid molecule encoding an enzyme involved in lignan biosynthetic pathway, useful for producing large amounts of lignans

XX

PS Claim 9; Page 55-56; 66pp; English.

XX

CC The present sequence encodes a secoisolariciresinol dehydrogenase protein. The enzyme is involved in the lignan biosynthetic pathway. The secoisolariciresinol dehydrogenase proteins have a molecular weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase nucleic acids are used for the recombinant expression of the enzymes. It is also used to obtain expression or enhanced expression of secoisolariciresinol dehydrogenase in plants or to alter lignan biosynthesis. The enzyme is used for production of the pharmacologically active lignan, matairesinol. The secoisolariciresinol dehydrogenase proteins and nucleic acids can be utilized to: elevate or otherwise alter the levels of health-protecting lignans, including phytoestrogens such as enterolactone and enterodiol, in plant species, including vegetables, grains and fruits and to food items incorporating material derived from such genetically altered plants; genetically alter plant species to provide an abundant, natural supply of lignans useful for a variety of purposes, for example as neutriceuticals and dietary supplements; to genetically alter living organisms to produce an abundant supply of optically pure lignans having desirable biological properties, for example (-)-trachelogenin which possesses antiviral properties, and (-)-podophyllotoxin.

XX

SQ Sequence 828 BP; 235 A; 172 C; 196 G; 225 T; 0 other;

Query Match 11.1%; Score 91; DB 21; Length 828;

Best Local Similarity 100.0%; Pred. No. 4.1e-36;

Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CACAGCAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGTCACAGATGA 143

|||||

Db 93 CACAGCAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGTCACAGATGA 152

QY 144 ATTAGTCACTCAGTTGTCGAGGCCATTGGC 174

|||||

Db 153 ATTAGTCACTCAGTTGTCGAGGCCATTGGC 183

RESULT 6

AAAL38432

ID AAL38432 standard; DNA; 4505 BP.

XX

AC AAL38432;

XX

DT 15-AUG-2002 (first entry)

XX

DE Plasmid pAPI249 encoding secoisolariciresinol dehydrogenase.

XX

KW Guaiacyl (G)-lignan; monocotyledon plant; rice; food additive; seed-specific transcriptional regulatory region; dehydrogenase; gene; ds.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT CDS 953..1783

FT /*tag= a

FT /product= "Protein of secoisolariciresinol dehydrogenase"

FT /note= b

XX

PN WO200220548-A1.

XX

PD 14-MAR-2002.

XX

PF 04-SEP-2001; 2001WO-US27500.

XX

PR 07-SEP-2000; 2000US-230632P.

XX

PA (UNIW) UNIV WASHINGTON STATE RES FOUND.

XX

PI Lewis NG, Davin LB, Huang N;

XX

DR WPI; 2002-425767/45.

XX

DR P-PSDB; AAO21494, AAO21498.

XX

PT Increasing guaiacyl-lignan content in monocotyledon plants, by transforming plant with chimeric gene construct having seed-specific transcriptional regulator linked to gene encoding protein involved in G-lignan formation

XX

PS Claim 9; Fig 10; 136pp; English.

XX

CC The invention relates to a method for increasing the guaiacyl (G)-lignan content in seeds of a monocotyledon plant, comprising selecting at least one protein or enzyme integral to the pathway leading to G-lignan formation, stably transforming a monocotyledon plant with chimeric gene (CG) constructs having a seed-specific transcriptional regulatory region operably linked to a nucleic acid sequence encoding the enzyme. The method of the invention is useful for stably transforming a monocotyledonous plant (e.g. rice) with CG constructs resulting in increased expression of the genes encoded by CG constructs. The G-lignan enriched seed composition is useful as a food additive. This polynucleotide sequence represents plasmid pAPI249 encoding secoisolariciresinol dehydrogenase relating to the invention.

XX

SQ Sequence 4505 BP; 1166 A; 1096 C; 1099 G; 1144 T; 0 other;

Query Match 11.1%; Score 91; DB 24; Length 4505;

Best Local Similarity 100.0%; Pred. No. 4.2e-36;

Matches	91;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	84	CACAGCAAAACTCTTCTCCCAACATGGAGGCCAAAAGTTGCGATTGTCGTGCATGATGA	143						
Db	1045	CACAGCAAAACTCTTCTCCCAACATGGAGGCCAAAAGTTGCGATTGTCGTGCATGATGA	1104						
QY	144	ATTAGTCACTCAGTTGTCGAGGCCATTGGC	174						
Db	1105	ATTAGTCACTCAGTTGTCGAGGCCATTGGC	1135						
RESULT 7									
ID	AAZ45733								
XX	AAZ45733 standard; cDNA; 831 BP.								
XX	AAZ45733;								
DT	06-APR-2000 (first entry)								
XX	Secoisolaricresinol dehydrogenase cDNA clone SMDEHY511.								
DE	XX								
KW	Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;								
KW	lignan; matairesinol; health-protecting lignan; phytoestrogen;								
KW	enterolactone; enterodiol; nutraceutical; dietary supplement;								
KW	(-)atrachelogenin; antiviral; (-)-podophyllotoxin; ss.								
OS	Forsythia intermedia.								
XX	XX								
FH	Key	Location/Qualifiers							
FT	CDS	1..831							
FT	/tag=	a							
FT	/product=	"secoisolaricresinol dehydrogenase"							
FT	/note=	"no termination codon given"							
XX	XX								
PN	WO9955846-Al.								
XX	XX								
PD	04-NOV-1999.								
XX	XX								
PF	23-APR-1999; 99WO-US08975.								
PR	24-APR-1998; 98US-0082977.								
XX	{UNIW } UNIV WASHINGTON STATE RES FOUND.								
PA	Xia Z, Costa MA, Davin LB, Lewis NG;								
PI	WPI; 2000-126356/11.								
DR	P-PSDB; AAY54415.								
XX	New nucleic acid molecule encoding an enzyme involved in lignan								
PT	biosynthetic pathway, useful for producing large amounts of lignans								
XX	-								
PS	Claim 9; Page 52-54; 66pp; English.								
CC	The present sequence encodes a secoisolaricresinol dehydrogenase								
CC	protein. The enzyme is involved in the lignan biosynthetic pathway.								
CC	The secoisolaricresinol dehydrogenase proteins have a molecular								
CC	weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require								
CC	NAD or NADP as a cofactor. The secoisolaricresinol dehydrogenase								
CC	nucleic acids are used for the recombinant expression of the enzymes.								
CC	It is also used to obtain expression or enhanced expression of								
CC	secoisolaricresinol dehydrogenase in plants or to alter lignan								
CC	biosynthesis. The enzyme is used for production of the pharmacologically								
CC	active lignan, matairesinol. The secoisolaricresinol dehydrogenase								
CC	proteins and nucleic acids can be utilized to: elevate or otherwise								
CC	alter the levels of health-protecting lignans, including phytoestrogens								
CC	such as enterolactone and enterodiol, in plant species, including								
CC	vegetables, grains and fruits and to food items incorporating material								
CC	derived from such genetically altered plants; genetically alter plant								
CC	species to provide an abundant, natural supply of lignans useful for								
CC	a variety of purposes, for example as nutraceuticals and dietary								
CC	supplements; to genetically alter living organisms to produce an								
CC	abundant supply of optically pure lignans having desirable biological								

CC	properties, for example (-)-trachelogenin which possesses antiviral
XX	properties, and (-)-podophylotoxin.
XX	
SQ	Sequence 831 BP; 244 A; 181 C; 195 G; 211 T; 0 other;
	Query Match 10.6%; Score 87; DB 21; Length 831;
	Best Local Similarity 100.0%; Pred. No. 4.6e-34;
	Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	84 CACAGCAAAACTCTCTCCCAACATGAGGCCAAAGTTGCCATTGCTGATGTCCCAAGATGA 143
Db	96 CACAGCAAAACTCTCTCCCAACATGAGGCCAAAGTTGCCATTGCTGATGTCCCAAGATGA 155
QY	144 ATTAGTCACTCAGTTGTCGAGGCCAT 170
Db	156 ATTAGTCACTCAGTTGTCGAGGCCAT 182
RESULT 8	
AAZ45744	
ID	AAZ45744 standard; DNA; 33 BP.
XX	
XX	AAZ45744;
XX	
DT	06-APR-2000 (first entry)
XX	
DE	PCR primer DEHY130NTERM for secoisolariciresinol dehydrogenase cDNA.
XX	
KW	Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
KW	lignan; matairesinol; health-protecting lignan; phytoestrogen;
KW	enterolactone; enterodiol; neutriceutical; dietary supplement;
KW	(-)-trachelogenin; antiviral; (-)-podophylotoxin; PCR primer; ss.
OS	Forsythia intermedia.
XX	
PN	W09955846-A1.
XX	
PD	04-NOV-1999.
XX	
PF	23-APR-1999; 99WO-US08975.
XX	
PR	24-APR-1998; 98US-0082977.
XX	
XX	(UNIW) UNIV WASHINGTON STATE RES FOUND.
XX	
PI	Xia Z, Costa MA, Davin LB, Lewis NG;
XX	
DR	WPI; 2000-126356/11.
XX	
PT	New nucleic acid molecule encoding an enzyme involved in lignan
PT	biosynthetic pathway, useful for producing large amounts of lignans
XX	
PS	Example 2; Page 63; 66pp; English

PCR primers AAZ45744-45 were used to amplify the secoisolariciresinol dehydrogenase cDNA sequence. The enzyme is involved in the lignan biosynthetic pathway. The secoisolariciresinol dehydrogenase proteins have a molecular weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase nucleic acids are used for the recombinant expression of the enzymes. It is also used to obtain expression or enhanced expression of secoisolariciresinol dehydrogenase in plants or to alter lignan biosynthesis. The enzyme is used for production of the pharmacologically active lignan, matairesinol. The secoisolariciresinol dehydrogenase proteins and nucleic acids can be utilized to: elevate or otherwise alter the levels of health-protecting lignans, including phytoestrogens such as enterolactone and enterodiol, in plant species, including vegetables, grains and fruits and to food items incorporating material derived from such genetically altered plants; genetically alter plant species to provide an abundant, natural supply of lignans useful for a variety of purposes, for example as nutraceuticals and dietary supplements; to genetically alter living organisms to produce an abundant supply of optically pure lignans having desirable biological

CC properties, for example (-)-trachelogenin which possesses antiviral
CC properties, and (-)-podophyllotoxin.

SQ Sequence 33 BP; 11 A; 8 C; 7 G; 7 T; 0 other;

Query Match 3.5%; Score 29; DB 21; Length 33;

Best Local Similarity 100.0%; Pred. No. 0.00018;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTTCGAACTGCATTCGCAAGAAG 29

|||||

Db 5 ATGCAGCTTCGAACTGCATTCGCAAGAAG 33

RESULT 9

AA245741

ID AA245741 standard; DNA; 23 BP.

XX AC

XX AA245741;

XX XX

DT 06-APR-2000 (first entry)

XX DE

PCR primer DEHYNTERM1 for secoisolariciresinol dehydrogenase cDNA.

XX KW

Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;

lignan; matairesinol; health-protecting lignan; phytoestrogen;

KW enterolactone; enterodiol; neutraceutical; dietary supplement;

KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; PCR primer; ss.

XX OS

Forsythia intermedia.

XX PN

WO9955846-A1.

XX PD

04-NOV-1999.

XX PF

23-APR-1999; 99WO-US089975.

XX PR

24-APR-1999; 98US-0082977.

XX PA

(UNIW) UNIV WASHINGTON STATE RES FOUND.

XX PI

Xia Z, Costa MA, Davin LB, Lewis NG;

XX WPI; 2000-126356/11.

XX DR

New nucleic acid molecule encoding an enzyme involved in lignan

biosynthetic pathway, useful for producing large amounts of lignans

XX PT

Example 2; Page 60; 66pp; English.

XX CC

The present PCR primer was used to amplify the secoisolariciresinol

dehydrogenase cDNA sequence. The enzyme is involved in the lignan

biosynthetic pathway. The secoisolariciresinol dehydrogenase proteins

have a molecular weight of 27-31 kDa, an isoelectric point of 5.9-6.85,

and require NAD or NADP as a cofactor. The secoisolariciresinol

dehydrogenase nucleic acids are used for the recombinant expression of

the enzymes. It is also used to obtain expression or enhanced expression

of secoisolariciresinol dehydrogenase in plants or to alter lignan

biosynthesis. The enzyme is used for production of the pharmacologically

active lignan, matairesinol. The secoisolariciresinol dehydrogenase

proteins and nucleic acids can be utilized to: elevate or otherwise

alter the levels of health-protecting lignans, including phytoestrogens

such as enterolactone and enterodiol, in plant species, including

vegetables, grains and fruits and to food items incorporating material

derived from such genetically altered plants; genetically alter plant

species to provide an abundant, natural supply of lignans useful for

a variety of purposes, for example as neutraceuticals and dietary

supplements; to genetically alter living organisms to produce an

abundant supply of optically pure lignans having desirable biological

properties, and (-)-trachelogenin which possesses antiviral

properties, and (-)-podophyllotoxin.

XX SQ

Sequence 23 BP; 6 A; 7 C; 5 G; 5 T; 0 other;

Query Match 2.8%; Score 23; DB 21; Length 23;

Best Local Similarity 100.0%; Pred. No. 0.21;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CAGCTTCGAACTGCATTCGCAAG 26

|||||

Db 1 CAGCTTCGAACTGCATTCGCAAG 23

RESULT 10

AAL38441/c

ID AAL38441 standard; DNA; 21 BP.

XX AC

AAL38441;

XX XX

DT 15-AUG-2002 (first entry)

XX DE

Dehydrogenase detecting PCR primer #2.

XX KW

Guaiacyl (G)-lignan; monocotyledon plant; rice; food additive;

KW seed-specific transcriptional regulatory region; PCR; primer; ss;

KW dehydrogenase.

XX OS

Unidentified.

XX PN

WO200220548-A1.

XX XX

PD 14-MAR-2002.

XX XX

04-SEP-2001; 2001WO-US27500.

XX PF

07-SEP-2000; 2000US-230632P.

XX PR

(UNIW) UNIV WASHINGTON STATE RES FOUND.

XX PA

(PHYT-) APPLIED PHYTOLOGICS INC.

XX PI

Lewis NG, Davin LB, Huang N;

XX DR

WPI; 2002-425767/45.

XX CC

Increasing guaiacyl-lignan content in monocotyledon plants, by
transforming plant with chimeric gene construct having seed-specific
transcriptional regulator linked to gene encoding protein involved in
G-lignan formation

XX PS

Example 5; Page 76; 136pp; English.

XX CC

The invention relates to a method for increasing the guaiacyl (G)-lignan

content in seeds of a monocotyledon plant, comprising selecting at least

one protein or enzyme integral to the pathway leading to G-lignan

formation, stably transforming a monocotyledon plant with chimeric gene

(CG) constructs having a seed-specific transcriptional regulatory region

operably linked to a nucleic acid sequence encoding the enzyme. The

method of the invention is useful for stably transforming a

monocotyledonous plant (e.g. rice) with CG constructs resulting in

increased expression of the genes encoded by CG constructs. The G-lignan

enriched seed composition is useful as a food additive. This

polynucleotide sequence represents a dehydrogenase detecting PCR primer

relating to the invention.

XX SQ

Sequence 21 BP; 7 A; 8 C; 4 G; 2 T; 0 other;

Query Match 2.5%; Score 21; DB 24; Length 21;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GGCTGGTGGTCTTCACATGC 482

|||||

Db 21 GGCTGGTGGTCTTCACATGC 1

RESULT 11

```
AAL38451/C
ID AAL38451 standard; DNA; 21 BP.
XX
AC AAL38451;
XX
DI 15-AUG-2002 (first entry)
XX
DE Gtl dehydrogenase PCR primer #2.
XX
KW Guaiacyl (G)-lignan: monocotyledon plant; rice; food additive;
KW seed-specific transcriptional regulatory region; PCR; primer; ss.
XX
OS Unidentified.
XX
PN WO200220548-A1.
XX
PD 14-MAR-2002.
XX
PF 04-SEP-2001; 2001WO-US27500.
XX
PR 07-SEP-2000; 2000US-230632P.
XX
PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.
PA (PHYT-) APPLIED PHYTOLOGICS INC.
PI Lewis NG, Davin LB, Huang N;
XX
DR WPI; 2002-425767/45.
XX
PT Increasing guaiacyl-lignan content in monocotyledon plants, by
PT transforming plant with chimeric gene construct having seed-specific
PT transcriptional regulator linked to gene encoding protein involved in
PT G-lignan formation -
XX
PS Example 5; Page 70; 136pp; English.
XX
CC The invention relates to a method for increasing the guaiacyl (G)-lignan
CC content in seeds of a monocotyledon plant, comprising selecting at least
CC one protein or enzyme integral to the pathway leading to G-lignan
CC formation, stably transforming a monocotyledon plant with chimeric gene
CC (CG) constructs having a seed-specific transcriptional regulatory region
CC operably linked to a nucleic acid sequence encoding the enzyme. The
CC method of the invention is useful for stably transforming a
CC monocotyledonous plant (e.g. rice) with CG constructs resulting in
CC increased expression of the genes encoded by CG constructs. The G-lignan
CC increased seed composition is useful as a food additive. This
CC polynucleotide sequence represents a Gtl dehydrogenase PCR primer
CC relating to the invention.
XX
SQ Sequence 21 BP; 7 A; 8 C; 4 G; 2 T; 0 other;

Query Match 2.6%; Score 21; DB 24; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GGGTGGTGGTCTTCACATGC 482
Db 21 GGGTGGTGGTCTTCACATGC 1
|||||
RESULT 12
AAFL2470/C
ID AAFL2470 standard; cDNA; 703 BP.
XX
AC AAFL2470;
XX
DI 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:4993.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

culture condition; environmental stress; spore morphogenesis;
metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO ) NOVO NORDISK BIOTECH INC.
PA (NOVO ) NOVO NORDISK AS.
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2099-2100; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random DNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAFL1248 to AAFL1853 represents ESTs from
CC Fusarium venenatum; AAFL1248 to AAFL1853 represents ESTs from Aspergillus
CC niger; AAFL1854 to AAFL14878 represents ESTs from Aspergillus oryzae; and
CC AAFL14879 to AAFL15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 703 BP; 157 A; 208 C; 189 G; 143 T; 6 other;

Query Match 2.3%; Score 19; DB 21; Length 703;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TTGCCATTGCTGATGTCCA 137
Db 76 TTGCCATTGCTGATGTCCA 58
|||||
RESULT 13
AEQ68055
ID ABO68055 standard; DNA; 1029 BP.
XX
AC ABO68055;
XX
DI 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes EGD DNA sequence #179.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
```

```

KW infection; ds.
XX
OS Listeria monocytogenes EGD.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators
XX
PS Claim 16; SEQ ID 868; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1029 BP; 377 A; 164 C; 214 G; 274 T; 0 other;

Query Match 2.3%; Score 19; DB 24; Length 1029;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 617 TTAATAATGAAGAAGATT 635
Db 275 TTAATAATGAAGAAGATT 293

RESULT 14
ABQ69995
ID ABQ69995 standard; DNA; 1029 BP.
XX
AC ABQ69995;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes EGDe DNA sequence #207.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes EGDe.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.

```

```

PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators
XX
PS Claim 16; SEQ ID 2808; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1029 BP; 377 A; 164 C; 214 G; 274 T; 0 other;

Query Match 2.3%; Score 19; DB 24; Length 1029;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 617 TTAATAATGAAGAAGATT 635
Db 275 TTAATAATGAAGAAGATT 293

RESULT 15
ABL78811
ID ABL78811 standard; cDNA; 480 BP.
XX
AC ABL78811;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:1789.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US17756.
XX
PR 26-MAY-2000; 2000US-207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI; 2002-122075/16.
XX
PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide
XX
PS Claim 1; SEQ ID 1789; 489pp; English.
XX
CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour

```

CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
 CC (SI) from the 10912 nucleotide sequences as given in ABL77023 to
 CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
 CC population of (II), or antigen presenting cells that express (II).
 CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
 CC (SI) can be used for detecting ovarian cancer in a patient's biological
 CC sample preferably serum or ovarian tissue. The method comprises
 CC contacting a biological sample from a patient with (IV), detecting the
 CC amount of polynucleotide hybridising to (IV) and comparing the amount to
 CC a predetermined cutoff value and thereby detecting ovarian cancer in the
 CC patient, where the amount of polynucleotide hybridising to (IV) is
 CC detected preferably by polymerase chain reaction (PCR). (I) comprising
 CC (III) and/or (II) is useful for stimulating and/or expanding T cells
 CC specific for an ovarian tumour protein comprising contacting T cells
 CC with (III) or (II). (III) is useful in design and preparation of
 CC ribozyme molecules for inhibiting expression of the tumour polypeptides
 CC and proteins in tumour cells; and to isolate a full length gene from a
 CC suitable library e.g., a tumour cDNA library using well known
 CC techniques.

XX
 SQ Sequence 480 BP; 133 A; 98 C; 91 G; 157 T; 1 other;

Query Match 2.2%; Score 18; DB 24; Length 480;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GGAGGAGCCAGTGGAAATT 75
 Db 424 GGAGGAGCCAGTGGAAATT 441
 |||||

Search completed: February 18, 2003, 18:37:14
 Job time : 255 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 17:47:44 ; Search time 2348 Seconds
(without alignments)
10151.277 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgttccaataatccagattctt 819

Scoring table: **OLIGO_NUC**
Gapop 60.0, Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	647	79.0	831	8	AF352735	AF352735 Forsythia
2	22	2.7	15072	1	AE006989	AE006989 Mycobacte
3	22	2.7	38970	1	MYCY10G2	Z95339 Mycobacteri
c	21	2.6	102655	2	AC099466	AC099466 Rattus no
5	21	2.6	119807	2	AP003753	AP003753 Oryza sat
c	21	2.6	181600	2	AC015813	AC015813 Homo sapi
c	21	2.6	202361	2	AC094900	AC094900 Rattus no
8	20	2.4	963	3	AF218330	AF218330 Leontodon
c	20	2.4	26758	3	U23181	U23181 Caenorhabdi
10	20	2.4	52994	9	AL137004	AL137004 Human DNA
c	20	2.4	95315	2	AC098131	AC098131 Rattus no
c	12	2.4	150877	2	AC104680	AC104680 Sus scrof
13	20	2.4	176699	2	AC021284	AC021284 Homo sapi
14	20	2.4	180748	2	AC129570	AC129570 Mus muscu
15	20	2.4	181210	9	AC011742	AC011742 Homo sapi
16	20	2.4	187508	9	AC104078	AC104078 Homo sapi
c	18	2.4	210821	10	AL731814	AL731814 Mouse DNA
c	20	2.4	211672	2	AC124694	AC124694 Mus muscu
20	20	2.4	221617	2	AC122394	AC122394 Mus muscu
21	19	2.3	702	11	PM7G6G	AL685209 Penicilli
22	19	2.3	783	8	AF092601	AF092601 Munnozia
23	19	2.3	783	8	AF092602	AF092602 Liabum gl
24	19	2.3	783	8	AF092603	AF092603 Sinclai
25	19	2.3	906	8	AF218343	AF218343 Picris pa
26	19	2.3	915	8	AF218351	AF218351 Hieracium
27	19	2.3	939	8	AF218346	AF218346 Taraxacum
28	19	2.3	949	8	AF218336	AF218336 Pilosella
29	19	2.3	960	8	AF218325	AF218325 Tolpis vi
30	19	2.3	961	8	AF218327	AF218327 Tolpis az
31	19	2.3	961	8	AF218329	AF218329 Tolpis co
32	19	2.3	966	8	AF218350	AF218350 Reichardi
33	19	2.3	967	8	AF218328	AF218328 Tolpis fa
34	19	2.3	969	8	AF218352	AF218352 Urospermu
35	19	2.3	971	8	AF218344	AF218344 Hyoseris
36	19	2.3	973	8	AF218338	AF218338 Andryala
37	19	2.3	980	8	AF218342	AF218342 Tolpis ca
38	19	2.3	987	8	AF218326	AF218326 Tolpis ba
39	19	2.3	1026	8	AF405250	AF405250 Nicoletti
40	19	2.3	1026	8	AF405251	AF405251 Hydropect
41	19	2.3	1026	8	AF405252	AF405252 Adenopapp
42	19	2.3	1026	8	AF405254	AF405254 Chrysacti
43	19	2.3	1026	8	AF405256	AF405256 Boeberoid
44	19	2.3	1026	8	AF405257	AF405257 Dysodiops
45	19	2.3	1026	8	AF405258	AF405258 Dysodia

ALIGNMENTS

RESULT 1
AF352735

LOCUS

DEFINITION

Forsythia x intermedia stem secoisolariciresinol dehydrogenase

AF352735 mRNA linear PLN 23-APR-2001

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Forsythia x intermedia

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Lamiales; Oleaceae; Forsythia.

1 (bases 1 to 831)

Xia,Z.Q., Costa,M.A., Pelissier,H.C., Davin,L.B. and Lewis,N.G.

	gene	TDPKYLPAAATERAEAGCTLRGSAELRAALDETTAGLRARPKRASA" 8849.. 9730 /gene="Mti1075"
	Query Match	2.7%; Score 22; DB 1; Length 15072;
	Best Local Similarity	100.0%; Pred. No. 1;
	Matches	22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	683	ATGTGGAGGATTGCCAATGC 704 Db 4151 ATGTGGAGGATTGCCAATGC 4172
RESULT 3		
MTCY10G2		
LOCUS		MTCY10G2 38970 bp DNA linear BCT 03-AUG-2001
DEFINITION		Mycobacterium tuberculosis H37Rv complete genome; segment 47/162.
ACCESSION		Z92539 AL123456
VERSION		29539.1 GI:3261714
KEYWORDS		.
SOURCE		Mycobacterium tuberculosis H37Rv.
ORGANISM		Mycobacterium tuberculosis H37Rv
REFERENCE		Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 38970) Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Ges S., Barry III R.C.E., Tekala F., Badcock K., Basham B., Brown D., Chillingworth T., Connor R., Davies K., Devlin K., Felwell T., Gentles S., Hamlin N., Harper J., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squires K., Sulston J.E., Taylor K., Whitehead S. and Barrell B.G.
TITLE		Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
JOURNAL		Nature 393 (6685), 537-544 (1998)
MEDLINE		98295987
PUBMED		9634230
REFERENCE		2 (bases 1 to 38970) Parkhill J.
AUTHORS		Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Molculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
JOURNAL		On Jun 27, 1998 this sequence version replaced gi:1869979.
COMMENT		Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/M.tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. Location/Qualifiers 1..38970 /organism="Mycobacterium tuberculosis H37Rv" /strain="H37Rv" /db_xref="taxon:83332" <1..38452 /organism="Mycobacterium tuberculosis H37Rv" /strain="H37Rv"
FEATURES	source	/db_xref="taxon:83332" /clone="Y10G2" complement(51.. 626) /gene="pth" complement(51.. 626) /gene="pth" /note="RV1014c, (MTCY10G2.35), pth, len: 191. Probable peptidyl-trna hydrolase, similar to eg PTH_ECOLI_P23932 peptidyl-trna hydrolase (ec 3.1.1.29) (194 aa), fasta scores, opt: 472, E(): 2.3e-25, (39.6% identity in 187 aa overlap)" /codon_start=1 /transl_table=11 /product="pth" protein_id="CAB06865.1" /db_xref="GI:1870014" /db_xref="SWISS-PROT:P96386" translation="MAEPLLVVGLGNPCANYARHNLFGVVADLLAARLGAKFKAKH RSGAVATGRSAGRSILVLAKPYCMNESGROIGPLAKFYSVPANITIVIHDDLLEFG RIRLKIGGEGGHNGLSRVAALGTKDFORVRIGRGPRPGKDKPAAFVLENFTFAERA EVPTICEAADATELLIEQGMEPAQRVHAM" complement(639.. 1286) /gene="rplY" complement(639.. 1286) /note="RV1015c, (MTCY10G2.34), rply, len: 215. Probable 50S ribosomal protein 12S similar to RL25_ECOLI_P02426 50S ribosomal protein 12S. escherichi (94 aa), fasta scores, opt:182, E(): 2.5e-05, (38.4% identity in 86 aa overlap) and tOCT-BACSU PI4194 general stress protein etc. bacillus subtilis (203 aa), fasta scores, opt: 260, E(): 1.4e-09, (28.4%identity in 201 aa overlap)" /codon_start=1 /transl_table=11 /product="rplY" protein_id="CAB06864.1" /db_xref="GI:1870013" /db_xref="SWISS-PROT:P96385" translation="MAKSASNLQRTVTETKGASRRARRAGKIIPAVLYHGHAEPQH LELPHDYAAVLRHSGTNVLTLDIAGKEQLALKHIIPIRTIQHADLVLVRGE KVVEVSVVVEGGAGPDITVTQETNSIEIAEALSIPQETVIEGAPEQTQLTAQOI ALPAGVSLISDPLLNVNVPKAPTAEELLEGEVAGAEAEAEAVEAGEAAGESE" complement(1503.. 2183) /gene="lpqT" complement(1503.. 2183) /note="RV1016c, (MTCY10G2.33), len: 226. Probable lipoprotein, similar to several M. tuberculosis hyp[othetical proteins e.g. YOH3_MYCTU_P71697_Cy21d4_03c (310 aa), fasta scores: opt: 329 z-score: 442.0 E(): 2e-17, 32.3% identity in 229 aa overlap: contains PS00013 Prokaryotic membrane lipoproteinlipid attachment site" /codon_start=1 /transl_table=11 /product="lpqT" protein_id="CAB06863.1" /db_xref="GI:1870012" /db_xref="SWISS-PROT:P96384" translation="MAGSRFCODSVRPDLAVAVATLMSAVACGPKSPDFQSILSTS PTTSAYSTTEPVPLMKVILESGVTGEPAPSSLDTLTVSIPTPGWAPMNPNPTF NEMTKARGSYETAMLAVFKHGFDDIAELKHGTADARLUSTNTFLSDSTADFNGFF SMHQSYDLHGRLHTWNRIVFPTGPAPPKORYLVOLLTTISLANEAVKHASDIIEAI AGFVVAAK" complement(2094.. 2126) /gene="lpqT" /note="FS00013 prokaryotic membrane lipoprotein lipid attachment site" attachment site" complement(2219.. 3199) /gene="prsa" complement(2219.. 3199) /note="RV1017c, (MTCY10G2.32), prsa, len: 326. Probale ribose-phosphate pyrophosphokinase, similar to eg
FEATURES	source	/db_xref="taxon:83332" /clone="Y10G2" complement(51.. 626) /gene="pth" complement(51.. 626) /gene="pth" /note="RV1014c, (MTCY10G2.35), pth, len: 191. Probable peptidyl-trna hydrolase, similar to eg PTH_ECOLI_P23932 peptidyl-trna hydrolase (ec 3.1.1.29) (194 aa), fasta scores, opt: 472, E(): 2.3e-25, (39.6% identity in 187 aa overlap)" /codon_start=1 /transl_table=11 /product="pth" protein_id="CAB06865.1" /db_xref="GI:1870014" /db_xref="SWISS-PROT:P96386" translation="MAEPLLVVGLGNPCANYARHNLFGVVADLLAARLGAKFKAKH RSGAVATGRSAGRSILVLAKPYCMNESGROIGPLAKFYSVPANITIVIHDDLLEFG RIRLKIGGEGGHNGLSRVAALGTKDFORVRIGRGPRPGKDKPAAFVLENFTFAERA EVPTICEAADATELLIEQGMEPAQRVHAM" complement(639.. 1286) /gene="rplY" complement(639.. 1286) /note="RV1015c, (MTCY10G2.34), rply, len: 215. Probable 50S ribosomal protein 12S similar to RL25_ECOLI_P02426 50S ribosomal protein 12S. escherichi (94 aa), fasta scores, opt:182, E(): 2.5e-05, (38.4% identity in 86 aa overlap) and tOCT-BACSU PI4194 general stress protein etc. bacillus subtilis (203 aa), fasta scores, opt: 260, E(): 1.4e-09, (28.4%identity in 201 aa overlap)" /codon_start=1 /transl_table=11 /product="rplY" protein_id="CAB06864.1" /db_xref="GI:1870013" /db_xref="SWISS-PROT:P96385" translation="MAKSASNLQRTVTETKGASRRARRAGKIIPAVLYHGHAEPQH LELPHDYAAVLRHSGTNVLTLDIAGKEQLALKHIIPIRTIQHADLVLVRGE KVVEVSVVVEGGAGPDITVTQETNSIEIAEALSIPQETVIEGAPEQTQLTAQOI ALPAGVSLISDPLLNVNVPKAPTAEELLEGEVAGAEAEAEAVEAGEAAGESE" complement(1503.. 2183) /gene="lpqT" complement(1503.. 2183) /note="RV1016c, (MTCY10G2.33), len: 226. Probable lipoprotein, similar to several M. tuberculosis hyp[othetical proteins e.g. YOH3_MYCTU_P71697_Cy21d4_03c (310 aa), fasta scores: opt: 329 z-score: 442.0 E(): 2e-17, 32.3% identity in 229 aa overlap: contains PS00013 Prokaryotic membrane lipoproteinlipid attachment site" /codon_start=1 /transl_table=11 /product="lpqT" protein_id="CAB06863.1" /db_xref="GI:1870012" /db_xref="SWISS-PROT:P96384" translation="MAGSRFCODSVRPDLAVAVATLMSAVACGPKSPDFQSILSTS PTTSAYSTTEPVPLMKVILESGVTGEPAPSSLDTLTVSIPTPGWAPMNPNPTF NEMTKARGSYETAMLAVFKHGFDDIAELKHGTADARLUSTNTFLSDSTADFNGFF SMHQSYDLHGRLHTWNRIVFPTGPAPPKORYLVOLLTTISLANEAVKHASDIIEAI AGFVVAAK" complement(2094.. 2126) /gene="lpqT" /note="FS00013 prokaryotic membrane lipoprotein lipid attachment site" attachment site" complement(2219.. 3199) /gene="prsa" complement(2219.. 3199) /note="RV1017c, (MTCY10G2.32), prsa, len: 326. Probale ribose-phosphate pyrophosphokinase, similar to eg
FEATURES	source	/db_xref="taxon:83332" /clone="Y10G2" complement(51.. 626) /gene="pth"<

KPRS_ECOLI P08330 ribose-phosphate pyrophosphokinase (314 aa), fasta scores, opt: 826, E(): 0, (43.8% identity in 317 aa overlap); contains PS00103 Purine/pyrimidine phosphoribosyl transferases signature; contains PS00144 asparaginase / glutaminase active site signature 1"

```
/codon_start=1
/transl_table=11
/product="prsa"
/protein_id="CAB06862.1"
/db_xref="GI:1870011"
/db_xref="SPTREMBL:P96383"
/translation="MSHWDTNRKMLFAGRAHPELAQVAKELDVHVTSDAREFA
NGEIVFHVSGDCAPVQLQSPAPVNRWLMQELIMIDALKRGSAKITAVMPYPY
ARQDKHGRREPISARLADLTKTAGADRIIVTLDTIQGFFDGPVHRGQNLIT
GYRDNYPDGNVVPSPDSGRVRIAERKADALGGVPLAFIKHTRPVRPNVSVNRVV
GDVAGRTCLVDDMDTGTGTAGAVALLHNDGAGVVIATATHGVLSDDPAARLASCGA
REVIVNTLPIGEDKREFQLVLSIAPLASTIRAVFVSGVTGLFDGDA"
/complement(2480..2506)
/feature
  misc_feature
```

misc_feature

```
/note="PS00144 Asparaginase / glutaminase active site
signature 1"
/complement(2486..2524)
/feature
  misc_feature
  gene
  CDS
```

misc_feature

```
/note="PS00103 Purine/pyrimidine phosphoribosyl
transferases signature"
/complement(3291..4778)
/feature
  gene
  CDS
```

gene

CDS

```
/note="rv1018c, (MTCY10G2.31), glmU, len: 495. Probable
udp-n-acetylglucosamine pyrophosphorylase, similar to eg
GCAD_BACSU P14192 udp-n-acetylglucosamine
pyrophosphorylase (EC2.7.7.23) (456 aa), fasta scores,
opt: 1150, E(): 0, (40.0%identity in 453 aa overlap)"
/codon_start=1
/transl_table=11
/product="glmU"
/protein_id="CAB06861.1"
/db_xref="GI:1870010"
/db_xref="SPTREMBL:P96382"
/translation="MTFPGDTAVLVLAAGPGTRMSRTPKVLHTLAGRSMLSHLVHAI
AKLAPQRLIVGLGHQRIAPLVLGELADTLGRITDVALQDPLGTGTHAVLGLSLALPD
DVAGNVVTSPTDPLDADTLADLIATHRAVSAVTLTTLDLDPFGRLRLTODHE
VMAIYEQDTPSPQREINAGVYAFDIALRSALSRLSNNAQOELLTDVIALIR
SDQCVHSHVDDSLAVGVNVRVQLAELSELNRVVAAGVACVTVDPDATTWIDV
DVTIGRDVVIHPTGOLLGTOIGRCVVGPDITLTDVAVGDGASVVRTHSGSSSITGDG
RAVGPFTYLRPCTALGADGKLGAFVEVNSITGCTKYPHLTYVDADIGYSNIGAS
SVFVNYDTSKRRTVGSVHRTGSDTMFVAPVTIGDGTAGTVVREDVPPGALVAS
AGPQRNIENVORRKPSPGAPAAQASKRASEMACQQTQPPDADQTP"
/complement(4794..4865)
/product="trNA-Gln"
/note="gInt, trNA-Gln, anticodon ttt, length = 72"
5033..5626
/feature
  gene
  CDS
```

tRNA

gene

CDS

```
/note="rv1019, (MTCY10G2.30c), len: 197. Probable
transcriptional regulator, similar to many members of the
tetR family e.g. MTCY7D11.18c, (34.4% identity in 189 aa
overlap). Helix turn helix motif from aa 27-48 (+5.42 SD)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1019"
/protein_id="CAB06860.1"
/db_xref="GI:1870009"
/db_xref="SPTREMBL:P96381"
/translation="MTGTERRHLQIGIARSLFAERGYDGTSEIEIAQRANYSKPVVYE
HFGKGEGYAVVVDREMSALLDGTISLTNNRSVRVERVALALTYVEERTDGRIM
IRDSPASISSTGYSLNDVAYSQVSSILAGDFARGLDPLAPLVAQALVGSVSMTAQ
```

Query Match

Best Local Similarity 2.7%; Score 22; DB 1; Length 38970;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 683 ATGTTGAGGATGTTGCCAATGC 704
|||||
Db 29694 ATGTTGAGGATGTTGCCAATGC 29715

RESULT 4

AC099466

LOCUS

DEFINITION

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

AC099466

```
COMMENT
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17974364.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTW
Center clone name: CH20-6319
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 53408 bases at least Q40
Consensus quality: 56431 bases at least Q30
Consensus quality: 58711 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 52 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
1 1125: contig of 1125 bp in length
1126 1225: gap of unknown length
1226 2263: contig of 1038 bp in length
2264 2363: gap of unknown length
2364 3677: contig of 1314 bp in length
3678 3777: gap of unknown length
3778 5022: contig of 1245 bp in length
5023 5122: gap of unknown length
5123 6434: contig of 1312 bp in length
6435 6534: gap of unknown length
6535 8059: contig of 1525 bp in length
8060 8159: gap of unknown length
8160 9232: contig of 1073 bp in length
9233 9332: gap of unknown length
9333 10695: contig of 1363 bp in length
10696 11806: contig of 1011 bp in length
11807 11906: gap of unknown length
11907 12975: contig of 1069 bp in length
12976 14214: contig of 1139 bp in length
14215 14314: gap of unknown length
14315 15501: contig of 1187 bp in length
15502 16620: contig of 1019 bp in length
16621 16720: gap of unknown length
16721 18074: contig of 1354 bp in length
18075 18374: gap of unknown length
18375 20019: contig of 1845 bp in length
20020 20119: gap of unknown length
20120 21674: contig of 1555 bp in length
21675 23256: contig of 1482 bp in length
23257 23356: gap of unknown length
23357 24782: contig of 1426 bp in length
24783 24882: gap of unknown length
24883 26781: contig of 1899 bp in length
26782 26881: gap of unknown length
26882 28181: contig of 1300 bp in length
28182 28281: gap of unknown length
28282 29326: contig of 1045 bp in length
29327 29426: gap of unknown length
29427 30677: contig of 1251 bp in length
30678 30777: gap of unknown length
30778 32054: contig of 1277 bp in length
32054 32054: contig of 1277 bp in length
32055 32154: gap of unknown length
32155 33520: contig of 1366 bp in length
33521 35221: gap of unknown length
35222 35765: contig of 2145 bp in length
35766 35865: gap of unknown length
35866 37338: contig of 1473 bp in length
37339 37438: gap of unknown length
37439 39292: contig of 1854 bp in length
39293 39382: gap of unknown length
39383 41781: contig of 2389 bp in length
41782 44331: contig of 2550 bp in length
44332 44531: gap of unknown length
44532 45988: contig of 1457 bp in length
45989 46088: gap of unknown length
46089 47495: gap of unknown length
47496 50198: contig of 2604 bp in length
50199 50298: gap of unknown length
50299 51745: contig of 1447 bp in length
51746 51845: gap of unknown length
51846 53391: contig of 1546 bp in length
53392 54988: contig of 1497 bp in length
54989 55088: gap of unknown length
55089 57068: contig of 1980 bp in length
57069 57168: gap of unknown length
57169 59336: contig of 2768 bp in length
59337 60037: gap of unknown length
60038 61882: contig of 1746 bp in length
61883 64005: contig of 2123 bp in length
64006 64105: gap of unknown length
64106 65295: contig of 1190 bp in length
65296 65395: gap of unknown length
65396 67377: contig of 1982 bp in length
67378 67477: gap of unknown length
67478 68685: contig of 1208 bp in length
68686 68785: gap of unknown length
68786 70813: contig of 2027 bp in length
70814 70912: gap of unknown length
70913 73644: contig of 2731 bp in length
73645 73743: gap of unknown length
73744 77075: contig of 3331 bp in length
77076 77175: gap of unknown length
77176 80666: contig of 3492 bp in length
80667 80766: gap of unknown length
80767 83466: contig of 2700 bp in length
83467 83566: gap of unknown length
83568 87526: contig of 3960 bp in length
87527 87626: gap of unknown length
87627 90412: contig of 2786 bp in length
90413 90512: gap of unknown length
90513 94867: contig of 4355 bp in length
94868 94967: gap of unknown length
94968 97861: contig of 2894 bp in length
97862 97961: gap of unknown length
97962 102655: contig of 4694 bp in length.
102655 102655: contig of 4694 bp in length.
FEATURES
Location/Qualifiers
1..102655
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
Query Match 2.6%; Score 21; DB 2; Length 102655;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 605 AATTTTCAGGATTAATAATG 625
|||||
DB 67018 AATTTTCAGGATTAATAATG 67038
|||||
RESULT 5
```

```

AP003753/c
LOCUS       AP003753               119807 bp    DNA        linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 7 clone
            OJ1339_B08, *** SEQUENCING IN PROGRESS ***
ACCESSION   AP003753
VERSION     AP003753.1  GI:14422471
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:OJ1339_B08.
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1
AUTHORS     Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE       Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1339_B08
JOURNAL     Published Only in Database (2001)
REFERENCE   2
AUTHORS     Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE       Direct Submission
JOURNAL     Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
            The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source           1..119807
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="7"
                     /clone="OJ1339_B08"
BASE COUNT  34992 a 23815 c 24131 g 36815 t      54 others
ORIGIN

Query Match      2.6%; Score 21; DB 2; Length 119807;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 AATTGGCATTAGGTTAATT 565
|||||
Db 18836 AATTGGCATTAGGTTAATT 18816

RESULT 6
AC015813
LOCUS       AC015813               181600 bp    DNA        linear   HTG 06-AUG-2002
DEFINITION  Homo sapiens clone RP11-159D12, *** SEQUENCING IN PROGRESS ***
ACCESSION   AC015813
VERSION     AC015813.11  GI:22122998
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Birren,B., Nusbaum,C. and Lander,E.
TITLE       1 (bases 1 to 181600)
JOURNAL     Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 181600)
            Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murray,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
            Phukhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zemlek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Aug 6, 2002 this sequence version replaced gi:21591957.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L398
            Center clone name: 159_D_12
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1 19146: contig of 19146 bp in length
            * 19147 19246: gap of 100 bp
            * 19247 181600: contig of 162354 bp in length.
            Location/Qualifiers
            1..181600
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Rp11-159D12"
            /clone_lib="RP11-11 Human Male BAC"
BASE COUNT  50395 a 41057 c 40282 g 49766 t      100 others
ORIGIN

```

Query Match	2.6%; Score 21; DB 2;	Length 181600;			
Best Local Similarity	100.0%; Pred. No. 3.3;				
Matches	21; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY 167	CCATTGGCAGCTTCCAAATTCCTCA 187				
Db 91881	CCATTGGCAGCTTCCAAATTCCTCA 91901				
RESULT 7					
AC094900/c					
LOCUS	AC094900	202361 bp	DNA	linear	HTG 22-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-6J7, *** SEQUENCING IN PROGRESS ***, 88 unordered pieces.				
ACCESSION	AC094900				
VERSION	AC094900.5	GI:21909431			
KEYWORDS	HTG; HTGS_PHASE1				
SOURCE	Rattus norvegicus				
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
REFERENCE	1 (bases 1 to 202361)				
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaral, H.C., Are,J.R., Ayele,M., Banks,T., Barbacia,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,T., Bowler,J., Bowie,S., Brieva,M., Brown,M., Brown,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Fieggs,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scher,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,K., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Woodson,C., Worley,K., Wu,C., Wu,Y., Wu,X., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,K.G., and Gibbs,R.				
	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 202361)				
AUTHORS	Worley,K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 202361)				
AUTHORS	Worley,K.C.				
Direct Submission					
Submitted (22-JUL-2002)	Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
On Jul 19, 2002	this sequence version replaced gi:21903566.				
-----	Genome Center				
Center:	Baylor College of Medicine				
Center code:	BCM				
Web site:	http://www.hgsc.bcm.tmc.edu/				
Contact:	hgsc-help@bcm.tmc.edu				
-----	Project Information				
Center project name:	GBRE				
Center clone name:	CH230-6J7				
-----	Summary Statistics				
Sequencing vector:	Plasmid;				
Chemistry:	Dye-terminator Big Dye; 100% of reads				
Assembly program:	Phrap; version 0.990329				
Consensus quality:	124551 bases at least Q40				
Consensus quality:	131182 bases at least Q30				
Consensus quality:	135595 bases at least Q20				

* NOTE:	Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).				
* NOTE:	This is a 'working draft' sequence. It currently consists of 88 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.				
1	1043:	contig of 1043 bp in length			
1044	1143:	gap of unknown length			
1144	2195:	contig of 1052 bp in length			
2196	2295:	gap of unknown length			
2296	3391:	contig of 1096 bp in length			
3392	3491:	gap of unknown length			
3492	4697:	contig of 1206 bp in length			
4698	4797:	gap of unknown length			
4798	5836:	contig of 1039 bp in length			
5837	5936:	gap of unknown length			
5937	7011:	contig of 1075 bp in length			
7012	7111:	gap of unknown length			
7112	8165:	contig of 1054 bp in length			
8166	8265:	gap of unknown length			
8266	9503:	contig of 1238 bp in length			
9504	9603:	gap of unknown length			
9604	10688:	contig of 1065 bp in length			
10689	10768:	gap of unknown length			
10769	12186:	contig of 1418 bp in length			
12187	12286:	gap of unknown length			
12287	13747:	contig of 1461 bp in length			
13748	13847:	gap of unknown length			
13848	15120:	contig of 1273 bp in length			
15121	15220:	gap of unknown length			
15221	16607:	contig of 1387 bp in length			
16608	16707:	gap of unknown length			
16708	18470:	contig of 1763 bp in length			
18471	18570:	gap of unknown length			
18571	19646:	contig of 1076 bp in length			
19647	19746:	gap of unknown length			
19747	21428:	contig of 1682 bp in length			
21429	21528:	gap of unknown length			
21529	22681:	contig of 1153 bp in length			
22682	22782:	gap of unknown length			
22783	24556:	contig of 1775 bp in length			
24557	24656:	gap of unknown length			
24657	26009:	contig of 1353 bp in length			
26010	26109:	gap of unknown length			
26110	27291:	contig of 1182 bp in length			
27292	27391:	gap of unknown length			
27392	28689:	contig of 1298 bp in length			
28690	28789:	gap of unknown length			

* 28790 29866: contig of 1077 bp in length
* 29867 29866: gap of unknown length
* 29967 31021: contig of 1055 bp in length
* 31021 31121: gap of unknown length
* 31121 32478: contig of 1357 bp in length
* 32478 32578: gap of unknown length
* 32578 33982: contig of 1404 bp in length
* 33982 34082: gap of unknown length
* 34082 35337: contig of 1255 bp in length
* 35337 35437: gap of unknown length
* 35437 36895: contig of 1458 bp in length
* 36895 36995: gap of unknown length
* 36995 38416: contig of 1421 bp in length
* 38416 38516: gap of unknown length
* 38516 40009: contig of 1493 bp in length
* 40009 40109: gap of unknown length
* 40109 41677: contig of 1568 bp in length
* 41677 43476: contig of 1699 bp in length
* 43476 43576: gap of unknown length
* 43576 44726: contig of 1150 bp in length
* 44726 44826: gap of unknown length
* 44826 46863: contig of 2037 bp in length
* 46863 47969: contig of 1006 bp in length
* 47969 48069: gap of unknown length
* 48069 49211: contig of 1142 bp in length
* 49211 49311: gap of unknown length
* 49311 50769: contig of 1458 bp in length
* 50769 50869: gap of unknown length
* 50869 53404: contig of 2435 bp in length
* 53404 55313: contig of 1909 bp in length
* 55313 55413: gap of unknown length
* 55413 56546: contig of 1133 bp in length
* 56546 56846: gap of unknown length
* 56846 57656: contig of 1010 bp in length
* 57656 57756: gap of unknown length
* 57756 59322: contig of 1466 bp in length
* 59322 61550: contig of 2228 bp in length
* 61550 61850: gap of unknown length
* 61850 63277: contig of 1626 bp in length
* 63277 64933: contig of 1557 bp in length
* 64933 65033: gap of unknown length
* 65033 67183: contig of 2150 bp in length
* 67183 67283: gap of unknown length
* 67283 69479: contig of 2196 bp in length
* 69479 71824: contig of 2245 bp in length
* 71824 71924: gap of unknown length
* 71924 73839: contig of 1915 bp in length
* 73839 73939: gap of unknown length
* 73939 75368: contig of 1429 bp in length
* 75368 75469: gap of unknown length
* 75469 77661: contig of 2193 bp in length
* 77661 77761: gap of unknown length
* 77761 79659: contig of 1898 bp in length
* 79659 79759: gap of unknown length
* 79759 82318: contig of 2359 bp in length
* 82318 84219: gap of unknown length
* 84219 84321: contig of 1803 bp in length
* 84321 86104: gap of unknown length
* 86104 86432: contig of 1783 bp in length

Query Match 2.6%; Score 21; DB 2; Length 202361;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 AATTTTCAGGATTAATAATG 625
|||||
DB 96745 AATTTTCAGGATTAATAATG 96725

RESULT 8
LOCUS AF218330 963 bp DNA linear PLN 12-JUN-2001
DEFINITION. Leontodon saxatilis NADH dehydrogenase (ndhf) gene, partial cds; chloroplast gene for chloroplast product.
ACCESSION AF218330
VERSION AF218330.1 GI:10180318
KEYWORDS Leontodon saxatilis.
SOURCE Chloroplast Leontodon saxatilis
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Leontodon.
REFERENCE 1 (bases 1 to 963)
AUTHORS Park,S.-J., Korompai,E.J., Francisco-Ortega,J., Santos-Guerra,A. and Jansen,R.K.
TITLE Phylogenetic relationships of Tolpis (Asteraceae: Lactuceae) based on ndhf sequences
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 963)
AUTHORS Jansen,R.K. and Park,S.-J.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1999) Integrative Biology, University of Texas at Austin, 24th at Whitis, Austin, TX 78712, USA
FEATURES
source Location/Qualifiers
1..963
/organism="Leontodon saxatilis"
/organelle="plastid:chloroplast"
/db_xref="taxon:122540"
<1..963
/gene="ndhf"
/gene="ndhf"
/gene="ndhf"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase"
/protein_id="AAG14009.1"
/db_xref="GI:10180319"
/translation="NDSWLYSPFIIIGIAMSAGTAFYMCRIYLLTFEGHLMVHPQNY SGKNIPLYSISLMGEGSKISNKNFGLITLLMKKNSRPPSPFSNKVYKIDENVRNMI QPFLSIPHFQNTKTYSPYEDSNTMLFPILVLILFTLVGFLGIPFNQDGLSKWLT PSINLLHKNSNLLIDWYEFSDKAPFSVSIAGFIAFLPKPVYSSQNDGLFLNAFY KMGPNRIYDKIKNAIYDMSYRGVYDAFYQGFETAGIRRLADEAHFFDRRIIDAIPN GVGLSFFVAEVIKSVGGGRISSYLFFVFSVAIFLLIYFLNL"
BASE COUNT 296 a 130 c 153 g 384 t
ORIGIN
Query Match 2.4%; Score 20; DB 8; Length 963;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 616 ATTAATAATGAAGAGAAATT 635
|||||
DB 225 ATTAATAATGAAGAGAAATT 244
RESULT 9
LOCUS U23181/c 26758 bp DNA linear INV 29-MAY-2002
DEFINITION. Caenorhabditis elegans cosmid ZK84, complete sequence.
ACCESSION U23181
VERSION U23181.1 GI:726435
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 26758)
AUTHORS Waterston,R.
TITLE Genome sequence of the nematode C. elegans: a platform for

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT	investigating biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998) 95069613 9851916 2 (bases 1 to 26758) Kirsten.J. The sequence of C. elegans cosmid ZK84 Unpublished (2001) 3 (bases 1 to 26758) Waterston,R. Direct Submission Submitted (29-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA Submitted by: Genome Sequencing Center Department of Genetics, Washington University St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RQ, England email: rwnematode.wustl.edu and jes@sanger.ac.uk NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions. This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone. For a graphical representation of this cosmid sequence and its analysis see: http://www.wormbase.org/db/seq/sequence?name=ZK84;class=Sequence	gene	/codon_start=1 /product="C. elegans INS-5 protein (corresponding sequence ZK84.3)" /protein_id="AAC48209.1" /db_xref="GI:2914124" /translation="MHSIVALMLIGTILPIAALHOKHQGFTLSSDSTGNQPMDAISR ADPHTNYRSCALGSSRMSENGIDVAOKCCSDYIKELCCPFD" 4009..4401 /gene="ins-6" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.6;class= Sequence" join(4009..4267,4322..4401) /gene="ins-6" /note="Contains similarity to Pfam domain: PF00049 (Insulin), Score=15.3, E-value=0.0041, N=1" /codon_start=1 /product="C. elegans INS-6 protein (corresponding sequence ZK84.6)" /protein_id="AAC48208.1" /db_xref="GI:2914123" /translation="MNSVFTIIFVLCAQVAASFQSGPSMSEASMOLLRELOHN MMESARPMRPRARRVPAGETRACRGLISLVMAVGDLCNPQEGKDIAIECCGQCS DDYRSACCP" complement(11974..12278) /gene="ZK84.4" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.4;class= Sequence" complement(join(11974..12134,12221..12278)) /gene="ZK84.4" /codon_start=1 /product="Hypothetical protein ZK84.4" /protein_id="AAC48206.1" /db_xref="GI:726439" /translation="MHEGPLEIRFLKFNFPDFPCSKELQELQORKKRTCTTTEQFLI LKPLIFLKFFTLIKLNLIIPSIPHCCI" 18023..19333 /gene="ZK84.2" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.2;class= Sequence" join(18023..18170,18541..18831,19182..19333) /gene="ZK84.2" /note="coded for by the following C. elegans CDNAS: YK207el.5" /codon_start=1 /product="Hypothetical protein ZK84.2" /protein_id="AAC48205.1" /db_xref="GI:726437" /translation="MTEEDVLPKVVVLVLFVSGMICYDVAYTMNRYTNDPASPKAN LIFGWAIVSSVDHYLTNDVFTGCGVMLECLINFAIGLAFAARSHGILLFT SNVVLWKVILFESVFLQPGEGHLPMPNPNFVSKEMFIWIPGVVWVLPVLVLFAL WNKLALPKMSKIKWASNKRCGLDSFGDQKRIALE" complement(19772..20026) /gene="ZK84.5" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.5;class= Sequence" complement(19772..20026) /gene="ZK84.5" /codon_start=1 /product="Hypothetical protein ZK84.5" /protein_id="AAC48207.1" /db_xref="GI:726440" /translation="MSOMQTLNSKIRKGNPNKVVIQDNGHEKKKTAAPATPKTEVT APPTEGFKIKDEVEMLRAGFKNHGKAVKSGQGTMIHEL" 24170..26737 /gene="ZK84.1" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.1;class= Sequence" join(24170..24234,24288..24357,24411..24597,24654..26737)
FEATURES source	Coding sequences below are the result of integration and manual review of the following data : computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE.INDEX.html) and The C. elegans ORFeome cloning project (http://wofdb.dfci.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964). Location/Qualifiers 1..26758 /organism="Caenorhabditis elegans" /strain="Bristol N2" /db_xref="taxon:6239" /chromosome="II" /clone="ZK84" 2002..2352 /gene="ins-5" /note="for a graphical representation of this gene see: http://www.wormbase.org/db/seq/sequence?name=ZK84.3;class= Sequence" join(2002..2188,2267..2352) /gene="ins-5"	gene	
CDS			

/gene="zk84.1"
/note="final exon in repeat region; similar to long tandem repeat region of sialidase (SP:TCNA_TRYCR, P23253) and neurofilament H protein; coded for by the following C. elegans cDNAs: yk56a10.3, yk56a10.5, CESE54f"
/codon_start=1
/product="Hypochemical protein zk84.1"
/protein_id="AAC48204.1"
/db_xref="GI:726436"
/translation="MRFSGFALVAVVAVAAPLPVLEKIKRTLDFLAKKTIGEVAIN DTGTFPIKIDNDNLDIFKPRPEIKLQERDNDAAAANKSTDAAVNSASVDYGYG DCAVADVAPGPEPTAPAEAGAEVPTAVEGSGVSGIPDEVAITVAADSGSGD APASNSVETLITATEPALIPPAEPKRIIPPSGIVTSEQQGAGDAPEAPVVEE TPATPAAETPAPATEASNAVGTPEGVYDGTANAASVAPEAVETPAPAPAEETPA PATSEEPAPAAETPAPAAETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPA DSAASADTAAALVDTSSEHAAEETAPADIAATETTPAPVPAPVADAAAGYDPS SIPEETPAPAAETPAPAAEETPAPAAEETPAPETASAPDAAGGAPADVAA PADVATPAPETSSAAGSYDVPSEASEVTAPIVESNTEAPSDSAIPGPAASEPA PAPIEAPDAATLETAPAAEPAPAAEAGYDAPSVPEETPAPAPAADETAPAA PAAETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPA PAPAEEETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPA PQAEEETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPAPAAEETPAPAAEGAAA PAPIVSSGYDAQSTDTVVASSAPATSSDASGATNVEGQQAIFAIQARKKVRRAFFGVG GRA"

BASE COUNT 8374 a 5213 c 4710 g 8461 t
ORIGIN
Query Match 2.4%; Score 20; DB 3; Length 26758;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 622 AATGAAGAAGAAATTGAGAA 641
|||||
Db 22320 AATGAAGAAGAAATTGAGAA 22301

RESULT 10
AL137004
LOCUS
DEFINITION Human DNA sequence from clone RPl-216A11 on chromosome 6. Contains ESTs and GSSs, complete sequence.
ACCESSION AL137004
VERSION AL137004.7 GI:9581552
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 52994)
Tracey,A.
Direct Submission
Submitted (14-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 28, 2000 this sequence version replaced gi:9367914.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
RPl-216A11 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RPl-216A11 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RPl-216A11 is at 1 in this sequence. The true left end of clone RPl-135L22 is at 52895 in this sequence.

FEATURES

Location/Qualifiers
1..52994
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RPl-216A11"
/clone_lib="RPCI-1"
2076..2303
/note="MIR repeat: matches 8..242 of consensus"
2570..3386
/note="L1PAL6 repeat: matches 5332..6157 of consensus"
Complement(3036..3831)
/note="match: GSS: Em:AQ895852"
3511..3824
/note="AluY repeat: matches 1..310 of consensus"
4305..4444
/note="L1MB3 repeat: matches 5685..5828 of consensus"
4481..4790
/note="L1MB2 repeat: matches 5817..6125 of consensus"
4892..4963
/note="AluJ/FRAM repeat: matches 221..292 of consensus"
5109..5389
/note="AluX repeat: matches 18..299 of consensus"
Complement(5256..5528)
/note="match: GSS: Em:AQ333863"
6277..6587
/note="AluY repeat: matches 1..310 of consensus"
6679..7103
/note="MLTIC repeat: matches 22..464 of consensus"
8056..8313
/note="match: GSS: Em:AQ614111"
8547..8839
/note="AluJo repeat: matches 1..296 of consensus"
9425..9780
/note="L2 repeat: matches 1420..1815 of consensus"
9881..9953
/note="L2 repeat: matches 1686..1750 of consensus"
10202..10226
/note="MLTJ repeat: matches 111..135 of consensus"
10250..10468
/note="MER58A repeat: matches 1..222 of consensus"
10683..10718
/note="18 copies 2 mer tg 86% conserved"
10719..11012
/note="AluJo repeat: matches 1..285 of consensus"
11034..11342
/note="AluX repeat: matches 1..310 of consensus"
11753..12242
/note="match: GSS: Em:AQ504470"
12636..12944
/note="AluX repeat: matches 1..312 of consensus"
Complement(13241..13718)
/note="match: GSS: Em:AQ453999"
13488..13792
/note="AluX repeat: matches 1..305 of consensus"
13810..13979
/note="L1MC3 repeat: matches 7546..7739 of consensus"
14860..14935
/note="38 copies 2 mer tt 65% conserved"
15552..15716
/note="L1MB3 repeat: matches 6021..6184 of consensus"

Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usman, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 95315)
Worley, K.C.
Direct Submission
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 95315)
Worley, K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17975697.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGMA
Center clone name: CH230-15517
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 24703 bases at least Q40
Consensus quality: 27404 bases at least Q30
Consensus quality: 29760 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1228: contig of 1228 bp in length
* 1229 1328: gap of unknown length
* 1329 2380: contig of 1052 bp in length
* 2381 2480: gap of unknown length
* 2481 4025: contig of 1545 bp in length
* 4026 4125: gap of unknown length
* 4126 5273: contig of 1148 bp in length
* 5274 5373: gap of unknown length
* 5374 6792: contig of 1419 bp in length
* 6793 6892: gap of unknown length
* 6893 8263: contig of 1371 bp in length
* 8264 9584: gap of unknown length
* 9585 9684: contig of 1221 bp in length
* 9685 11200: contig of 1516 bp in length
* 11201 11300: gap of unknown length
* 11301 12388: contig of 1088 bp in length
* 12389 12489: gap of unknown length
* 12489 13555: contig of 1067 bp in length
* 13556 13656: gap of unknown length
* 13656 14809: contig of 1154 bp in length
* 14810 14909: gap of unknown length

16044: contig of 1135 bp in length
16144: gap of unknown length
16145 17476: contig of 1332 bp in length
17477 17576: gap of unknown length
17577 17689: contig of 1113 bp in length
17690 18789: gap of unknown length
18790 20014: contig of 1225 bp in length
20015 20114: gap of unknown length
20115 21754: contig of 1640 bp in length
21755 21854: gap of unknown length
21855 23225: contig of 1371 bp in length
23226 23225: gap of unknown length
23226 25003: contig of 1678 bp in length
25004 25103: gap of unknown length
25104 26236: contig of 1133 bp in length
26237 26336: gap of unknown length
26337 27859: contig of 1523 bp in length
27860 27959: gap of unknown length
27960 29863: contig of 1904 bp in length
29864 29964: contig of 1536 bp in length
29964 31500: gap of unknown length
31500 33222: contig of 1623 bp in length
33223 33222: gap of unknown length
33223 34463: contig of 1141 bp in length
34464 34563: gap of unknown length
34564 35733: contig of 1170 bp in length
35734 35833: gap of unknown length
35834 37005: contig of 1172 bp in length
37006 37105: gap of unknown length
37106 38260: contig of 1155 bp in length
38261 38360: gap of unknown length
38361 39391: contig of 1031 bp in length
39392 39491: gap of unknown length
39492 40765: contig of 1274 bp in length
40766 40865: gap of unknown length
40866 42016: contig of 1151 bp in length
42017 42116: gap of unknown length
42117 43321: contig of 1205 bp in length
43322 43421: gap of unknown length
43422 45129: contig of 1708 bp in length
45130 45229: gap of unknown length
45230 47027: contig of 1798 bp in length
47028 47127: gap of unknown length
47128 48225: contig of 1098 bp in length
48226 48325: gap of unknown length
48326 50301: contig of 1976 bp in length
50302 50401: gap of unknown length
50402 52152: contig of 1751 bp in length
52153 52252: gap of unknown length
52253 53970: contig of 1718 bp in length
53971 54070: gap of unknown length
54071 55268: contig of 1198 bp in length
55269 55368: gap of unknown length
55369 56556: contig of 1288 bp in length
56557 56756: gap of unknown length
56757 58254: contig of 1498 bp in length
58255 58354: gap of unknown length
58355 59445: contig of 1091 bp in length
59446 59545: gap of unknown length
59546 62045: contig of 2500 bp in length
62046 62145: gap of unknown length
62146 64566: contig of 2421 bp in length
64567 64666: gap of unknown length
64667 66958: contig of 2292 bp in length
66959 67058: gap of unknown length
67059 69063: contig of 2005 bp in length
69064 69163: gap of unknown length
69164 71523: contig of 2360 bp in length
71524 71623: gap of unknown length
71624 73080: contig of 1457 bp in length
73081 73180: gap of unknown length
73181 74950: contig of 1770 bp in length

```

* 74951 75050: gap of unknown length
* 75051 77416: contig of 2366 bp in length
* 77417 77516: gap of unknown length
* 77517 78338: contig of 1422 bp in length
* 78339 79038: gap of unknown length
* 79039 81846: contig of 2808 bp in length
* 81847 81946: gap of unknown length
* 81947 84023: contig of 2077 bp in length
* 84024 84123: gap of unknown length
* 84124 86119: contig of 1996 bp in length
* 86120 86219: gap of unknown length
* 86220 87407: contig of 1188 bp in length

Query Match      2.4%  Score 20; DB 2; Length 95315;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAGAAGCTAGAGGAAG 43
      |||||
Db 45574 AAGAAGCTAGAGGAAG 45555

RESULT 12
AC104680/c
LOCUS
DEFINITION
Sus scrofa clone RP44-434D6, WORKING DRAFT SEQUENCE, 5 unordered
pieces.
ACCESSION
AC104680.1 GI:17933850
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Sus scrofa.
SOURCE
Sus scrofa
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 150877)
AUTHORS
Akhter N., Ayele K., Beckstrom-Sternberg S.M., Benjamin B.,
Blakesley R.W., Bouffard G.G., Breen K., Brinkley C., Brooks S.,
Dietrich N.L., Granite S., Guan X., Gupta J., Haghighi P.,
Ho S.-I., Idoi J.R., Karlins E., Laric P., Lee Lin S.-Q.,
Legaspi R., Maduro O.L., Maduro V.B., Mashello C., Mastrian S.D.,
McGaspary J.C., McDowell J., Pearson R., Prasad A., Stantripo S.,
Thomas J.W., Thomas P.J., Touchman J.W., Tsurgeon C., Vogt J.L.,
Walker M.A., Wetherby K.D., Wiggins L., Young A., Zhang L.-H. and
Green E.D.
MISC
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 150877)
Green E.D.
Direct Submission
Submitted (19-DEC-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@hgrl.nih.gov
----- Project Information
Center project name: Cnn
Center clone name: 434D6
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 148948 bases at least Q40
Consensus quality: 149609 bases at least Q30
Consensus quality: 150008 bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 150477; sum-of-contigs
Quality coverage: 12.81x in Q20 bases; agarose-fp
Quality coverage: 12.51x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces

```

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4912: contig of 4912 bp in length
* 4913 5012: gap of unknown length
* 5013 30610: contig of 25598 bp in length
* 30611 30710: gap of unknown length
* 30711 49660: contig of 18950 bp in length
* 49661 49760: gap of unknown length
* 49761 86175: contig of 36415 bp in length
* 86176 86275: gap of unknown length
* 86276 150877: contig of 64602 bp in length.
FEATURES
source
1..150877
/organism="Sus scrofa"
/db_xref="taxon:9623"
/clone_lib="RP44-434D6"
/clone_lib="RP44"
1..4912
/note="assembly_fragment"
5013..30610
/note="assembly_fragment"
clone_end:77
vector_side:right
30711..49660
/note="assembly_fragment"
clone_end:SP6
vector_side:right
49761..86175
/note="assembly_fragment"
86276..150877
/note="assembly_fragment"
BASE COUNT 45369 a 29612 c 30757 g 44733 t 406 others
ORIGIN
Query Match      2.4%  Score 20; DB 2; Length 150877;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 784 AATCTGTAATCAAGTGTT 803
      |||||
Db 130771 AATCTGTAATCAAGTGTT 130752

RESULT 13
AC021284
LOCUS
DEFINITION
Homo sapiens clone RP11-21G24, WORKING DRAFT SEQUENCE, 18 unordered
pieces.
ACCESSION
AC021284.3 GI:7249031
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 176699)
AUTHORS
Birren B., Linton L., Nusbaum C. and Lander E.
TITLE
Homo sapiens, clone RP11-21G24
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 176699)
Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
Anderson S., Baldwin J., Barna N., Beckerly R., Beda F.,
Boguslavsky L., Boukhgalter B., Brown A., Burkett G., Castle A.,
Choepel Y., Collangelo M., Collins S., Collymore A., Cooke P.,
DeArelano K., Dewar K., Domino M., Doyle M., Fenesfor J.,
Ferreira P., FitzHugh W., Forrest C., Gage D., Galagan J.,
Gardyna S., Grant G., Hagos B., Heaford A., Horton L.,
Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,
Landers T., Lehoczy J., Levine R., Lieu C., Liu G., Locke K.,

```



```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 180748)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome, clone RP23-343H21
Unpublished
2 (bases 1 to 180748)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Canarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 180748)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Canarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 16, 2002 this sequence version replaced gi:22004418.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23770
Center clone name: 343_H_21
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179045 bases at least Q40
Consensus quality: 179810 bases at least Q30
Consensus quality: 180026 bases at least Q20
Insert size: 180200; agarose-fp
Insert size: 180148; sum-of-contigs
Quality coverage: 8.5 in Q20 bases; agarose-fp
Quality coverage: 8.6 in Q20 bases; sum-of-contigs

```

```

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 13353: contig of 13353 bp in length
* 13354 13453: gap of 100 bp
* 13454 15102: contig of 1649 bp in length
* 15103 15202: gap of 100 bp
* 15203 20389: contig of 5187 bp in length
* 20390 20489: gap of 100 bp
* 20490 34118: contig of 13629 bp in length
* 34119 34218: gap of 100 bp
* 34219 52624: contig of 18406 bp in length
* 52625 52724: gap of 100 bp
* 52725 137521: contig of 84797 bp in length
* 137522 137621: gap of 100 bp
* 137622 180748: contig of 43127 bp in length.
*
* Location/Qualifiers
* 1..180748
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /clone_lib="RPCI-23 Female Mouse BAC"
* 1..13353
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:left"
* misc_feature
* 13454..15102
* /note="assembly_fragment"
* 15203..20389
* /note="assembly_fragment"
* 20490..34118
* /note="assembly_fragment"
* 34219..52624
* /note="assembly_fragment"
* 52725..137521
* /note="assembly_fragment"
* 137622..180748
* /note="assembly_fragment"
* clone_end:T7
* vector_side:right"
*
* BASE COUNT 49564 a 38649 c 39535 g 52399 t 601 others
* ORIGIN
*
* Query Match 2.4%; Score 20; DB 2; Length 180748;
* Best Local Similarity 100.0%; Pred. No. 12;
* Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY 24 AAGAAGGCTAGAGGAAAG 43
* |||||
* Db 167531 AAGAAGGCTAGAGGAAAG 167550
*
* RESULT 15
* AC011742
* LOCUS Homo sapiens BAC clone RF11-214J9 from 2, complete sequence.
* DEFINITION AC011742
* ACCESSION AC011742
* VERSION AC011742.3 GI:9887800
* KEYWORDS HTG.
* SOURCE Homo sapiens.
* ORGANISM Homo sapiens
* Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
* Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
* REFERENCE 1 (bases 1 to 181210)
* AUTHORS Sulston,J.E. and Waterston,R.
* TITLE Toward a complete human genome sequence
* JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

```



```
repeat_region 9329. .9529
/rpt_family="L1"
repeat_region 9525. .9734
/rpt_family="L1"
repeat_region 9980. .10069
/rpt_family="MIR"
repeat_region 10070. .10375
/rpt_family="Alu"
repeat_region 10376. .10546
/rpt_family="MIR"
misc_feature 10882. .10914
note="similar to EST AAL48206 (NID:gl717613) zo53f11.sl"
repeat_region 10883. .11204
/rpt_family="Alu"
repeat_region 11209. .11479
/rpt_family="Alu"
repeat_region 11556. .11680
/rpt_family="Alu"
repeat_region 12186. .12495
/rpt_family="Alu"
repeat_region 12622. .12921
/rpt_family="Alu"
repeat_region 13091. .13564
/rpt_family="L1"
repeat_region 13593. .13844
/rpt_family="Alu"
repeat_region 13851. .13993
/rpt_family="L1"
repeat_region 14123. .14435
/rpt_family="Alu"
repeat_region 14554. .14850
/rpt_family="Alu"
repeat_region 15228. .15339
/rpt_family="ERV"
misc_feature 15397. .15989
note="similar to EST BE080155 (NID:g8470442)"
repeat_region 16224. .16562
/rpt_family="MALR"
repeat_region 16567. .16865
/rpt_family="Alu"
repeat_region 16869. .17004
```

Query Match 2.4% Score 20; DB 9; Length 181210;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 TCAACCTATGGAAACTGGA 269
|||||
Db 87347 TCAACCTATGGAAACTGGA 87366

Search completed: February 18, 2003, 19:33:48
Job time : 3385 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 02:02:38 ; Search time 251 Seconds
(without alignments)
7348.154 Million cell updates/sec

Title: US-09-673-918A-1
Perfect score: 819
Sequence: 1 atgcagcttcgaactgcatt.....tggccaataatccagattct 819

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	819	100.0	819	21	AAZ45730
2	816	99.6	816	21	AAZ45743
3	806.4	98.5	831	21	AAZ45731
4	736	89.9	828	21	AAZ45734
5	736	89.9	4505	24	AAZ38432
6	733.8	89.6	819	21	AAZ45732
7	712	86.9	831	21	AAZ45733
8	200.2	24.4	1120	21	AAZ40811
9	188.8	23.1	1025	16	AAQ94623

Secoisolariciresin
Secoisolariciresin
Secoisolariciresin
Secoisolariciresin
Plasmid pAPI249 en
Secoisolariciresin
Secoisolariciresin
Secoisolariciresin
Arabisopsis thalia
Horseradish peroxi

10	187	22.8	1136	21	AAZ38191	Arabidopsis thalia
11	185.4	22.6	852	21	AAZ43068 <th>Arabidopsis thalia</th>	Arabidopsis thalia
12	183.6	22.4	1138	21	AAZ45620 <th>Arabidopsis thalia</th>	Arabidopsis thalia
13	150.2	18.3	774	21	AAZ43159 <th>Arabidopsis thalia</th>	Arabidopsis thalia
c	149	18.2	833	24	ABN98918 <th>Arabidopsis thalia</th>	Arabidopsis thalia
15	148.4	18.1	774	21	AAZ50609 <th>Arabidopsis thalia</th>	Arabidopsis thalia
16	147.6	18.0	1002	21	AAZ40363 <th>Arabidopsis thalia</th>	Arabidopsis thalia
17	146.2	17.9	997	21	AAZ50610 <th>Arabidopsis thalia</th>	Arabidopsis thalia
18	145.4	17.8	861	21	AAZ50608 <th>Arabidopsis thalia</th>	Arabidopsis thalia
19	99.8	12.2	941	21	AAZ34703 <th>Arabidopsis thalia</th>	Arabidopsis thalia
20	98.2	12.0	863	21	AAZ98308 <th>A. thaliana gene i</th>	A. thaliana gene i
21	97.2	11.9	1236	16	AAZ45060 <th>Maize Ts2 cDNA nuc</th>	Maize Ts2 cDNA nuc
22	90	11.0	338	21	AAZ57005 <th>Pinus radiata tran</th>	Pinus radiata tran
23	88.4	10.8	421	21	AAZ56947 <th>Pinus radiata tran</th>	Pinus radiata tran
24	88.4	10.8	497	21	AAZ56998 <th>Pinus radiata tran</th>	Pinus radiata tran
25	86.8	10.6	440	21	AAZ56990 <th>Pinus radiata tran</th>	Pinus radiata tran
26	86.8	10.6	453	21	AAZ56982 <th>Pinus radiata tran</th>	Pinus radiata tran
27	86.8	10.6	475	21	AAZ57080 <th>Pinus radiata tran</th>	Pinus radiata tran
28	86.2	10.5	441	21	AAZ57063 <th>Pinus radiata tran</th>	Pinus radiata tran
29	85.8	10.5	323	21	AAZ57065 <th>Pinus radiata tran</th>	Pinus radiata tran
30	85.6	10.5	465	22	AAH87673 <th>Peppermint plant o</th>	Peppermint plant o
31	82.6	10.1	498	21	AAZ57104 <th>Pinus radiata tran</th>	Pinus radiata tran
32	80	9.8	1076	24	ABL56024 <th>Plodia interpuncte</th>	Plodia interpuncte
c	79.6	9.7	903	24	ABQ69008 <th>Listeria monocytog</th>	Listeria monocytog
34	79.6	9.7	8307	24	ABQ71075 <th>Listeria monocytog</th>	Listeria monocytog
c	78	9.5	6948	20	AAZ12941 <th>Enterococcus faeca</th>	Enterococcus faeca
36	73.2	8.9	810	24	ABN67211 <th>Streptococcus poly</th>	Streptococcus poly
37	71.8	8.8	311	21	AAZ56919 <th>Pinus radiata tran</th>	Pinus radiata tran
38	71.8	8.8	31930	24	ABQ67194 <th>Listeria innocua c</th>	Listeria innocua c
39	68.6	8.4	221	21	AAZ57012 <th>Pinus radiata tran</th>	Pinus radiata tran
40	67.4	8.2	7471	21	AAZ88789 <th>B. subtilis bac DN</th>	B. subtilis bac DN
41	67	8.2	221	21	AAZ56958 <th>Pinus radiata tran</th>	Pinus radiata tran
42	65.2	8.0	4281	24	ABA91715 <th>Commamonas sp. cycl</th>	Commamonas sp. cycl
43	62.6	7.6	306	21	AAZ56932 <th>Pinus radiata tran</th>	Pinus radiata tran
c	62.6	7.6	2944528	24	ABA03041 <th>Listeria monocytog</th>	Listeria monocytog
45	62.4	7.6	774	24	AAD28464 <th>Bacillus subtilis</th>	Bacillus subtilis

ALIGNMENTS

RESULT 1
AAZ45730
ID AAZ45730 standard; cDNA; 819 BP.
XX
AC AAZ45730;
XX
XX 06-APR-2000 (first entry)
DT
DE Secoisolariciresinol dehydrogenase cDNA clone DEHY133.
XX
KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
KW lignan; metairesinol; health-protecting lignan; phytoestrogen;
KW enterolactone; enterodiol; neutral; dietary supplement;
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
XX
OS Forsythia intermedia.
XX
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 1..819
FT /*tag= a
FT /product= "secoisolariciresinol dehydrogenase"
FT /note= "no termination codon given"

WO9955846-A1.
XX
PN
XX
PD 04-NOV-1999.
XX
PF 23-APR-1999; 99WO-US08975.
XX
PR 24-APR-1998; 98US-0082977.
XX
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.

XX Xia Z, Costa MA, Davin LB, Lewis NG;
 XX WPI: 2000-126356/11.
 DR P-PSDB: AAY54412.
 XX
 PT New nucleic acid molecule encoding an enzyme involved in lignan
 PT biosynthetic pathway, useful for producing large amounts of lignans
 XX
 PS Claim 9; Page 45-46; 66pp; English.
 XX
 CC The present sequence encodes a secoisolariciresinol dehydrogenase
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.
 CC The secoisolariciresinol dehydrogenase proteins have a molecular
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
 CC nucleic acids are used for the recombinant expression of the enzymes.
 CC It is also used to obtain expression or enhanced expression of
 CC secoisolariciresinol dehydrogenase in plants or to alter lignan
 CC biosynthesis. The enzyme is used for production of the pharmacologically
 CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase
 CC proteins and nucleic acids can be utilized to: elevate or otherwise
 CC alter the levels of health-protecting lignans, including phytoestrogens
 CC such as enterolactone and enterodiol, in plant species, including
 CC vegetables, grains and fruits and to food items incorporating material
 CC derived from such genetically altered plants; genetically alter plant
 CC species to provide an abundant, natural supply of lignans useful for
 CC a variety of purposes, for example as nutraceuticals and dietary
 CC supplements; to genetically alter living organisms to produce an
 CC abundant supply of optically pure lignans having desirable biological
 CC properties, for example (-)-trachelogenin which possesses antiviral
 CC properties, and (-)-podophyllotoxin.
 XX
 SQ Sequence 819 BP; 240 A; 166 C; 189 G; 224 T; 0 other;
 Query Match 100.0%; Score 819; DB 21; Length 819;
 Best Local Similarity 100.0%; Pred. No. 3.7e-230;
 Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 GCCTATTGCTTCAAAGCATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTC 540
 QY 541 GGACAAATTTGGCATTAGGGTTAAATGTTTCTCCTCTCGGGCTTCTACGGCTTTAGGC 600
 Db 541 GGACAAATTTGGCATTAGGGTTAAATGTTTCTCCTCTCGGGCTTCTACGGCTTTAGGC 600
 QY 601 AGAATATTTTCAGGGATTAAATAATGAGAGAAATTTGAGAACTTAATAAATTTCCGGGA 660
 Db 601 AGAATATTTTCAGGGATTAAATAATGAGAGAAATTTGAGAACTTAATAAATTTCCGGGA 660
 QY 661 AATTTGAAGTCCAAAATTTAATGTTGAGGATTTGCCAATGCCAGCTCTTTATCTGGCT 720
 Db 661 AATTTGAAGTCCAAAATTTAATGTTGAGGATTTGCCAATGCCAGCTCTTTATCTGGCT 720
 QY 721 AGTGATGAGGCAAAATACGTGAGTGGACAAATCTGTTCAATGATGAGGGTTTCAGCGTC 780
 Db 721 AGTGATGAGGCAAAATACGTGAGTGGACAAATCTGTTCAATGATGAGGGTTTCAGCGTC 780
 QY 781 TGCATTTCTGTAATCAAGTGTTCCTCAATATCCAGATTCT 819
 Db 781 TGCATTTCTGTAATCAAGTGTTCCTCAATATCCAGATTCT 819
 RESULT 2
 AAZ45743
 ID AAZ45743 standard; cDNA; 816 BP.
 AC AAZ45743;
 XX
 DT 06-APR-2000 (first entry)
 DE Secoisolariciresinol dehydrogenase cDNA clone DEH130.
 KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
 KW lignan; matairesinol; health-protecting lignan; phytoestrogen;
 KW enterolactone; enterodiol; nutraceutical; dietary supplement;
 KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
 OS Forsythia intermedia.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..816
 FT /*tag= a
 FT /product= "secoisolariciresinol dehydrogenase"
 FT /note= "no termination codon given; partial
 FT sequence"
 XX
 PN W09955846-A1.
 XX
 PD 04-NOV-1999.
 XX
 PF 23-APR-1999; 99WO-US08975.
 XX
 PR 24-APR-1998; 98US-0082977.
 XX
 PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
 XX
 PI Xia Z, Costa MA, Davin LB, Lewis NG;
 XX
 DR WPI: 2000-126356/11.
 DR P-PSDB: AAY54420.
 XX
 PT New nucleic acid molecule encoding an enzyme involved in lignan
 PT biosynthetic pathway, useful for producing large amounts of lignans
 XX
 PS Example 2; Page 61-62; 66pp; English.
 CC The present sequence encodes a secoisolariciresinol dehydrogenase
 CC protein. The enzyme is involved in the lignan biosynthetic pathway.
 CC The secoisolariciresinol dehydrogenase proteins have a molecular
 CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
 CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
 CC nucleic acids are used for the recombinant expression of the enzymes.

CC It is also used to obtain expression or enhanced expression of
CC secoisolariciresinol dehydrogenase in plants or to alter lignan
CC biosynthesis. The enzyme is used for production of the pharmacologically
CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase
CC proteins and nucleic acids can be utilized to: elevate or otherwise
CC alter the levels of health-protecting lignans, including phytoestrogens
CC such as enterolactone and enterodiol, in plant species, including
CC vegetables, grains and fruits and to food items incorporating material
CC derived from such genetically altered plants; genetically alter plant
CC species to provide an abundant, natural supply of lignans useful for
CC a variety of purposes, for example as nutraceuticals and dietary
CC supplements; to genetically alter living organisms to produce an
CC abundant supply of optically pure lignans having desirable biological
CC properties, for example (-)-trachelogenin which possesses antiviral
CC properties, and (-)-podophyllotoxin.
XX
SQ Sequence 816 BP; 239 A; 166 C; 188 G; 223 T; 0 other;

Query Match 99.8%; Score 816; DB 21; Length 816;
Best Local Similarity 100.0%; Pred. No. 2:8e-229;
Matches 816; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CAGCTCGAAGTTCGCAAGAGCTAGAGGAAAGTTGCCCTTATAACAGGAGGA 63
DB 1 CAGCTCGAAGTTCGCAAGAGCTAGAGGAAAGTTGCCCTTATAACAGGAGGA 60

QY 64 GCCAGTGGAAATTGGAGAAACACAGCAAAACTCTCTCCCAACATGGAGCAAAAGTTGCC 123
DB 61 GCCAGTGGAAATTGGAGAAACACAGCAAAACTCTCTCCCAACATGGAGCAAAAGTTGCC 120

QY 124 ATTGCTGATGTCGAAGATGAATTAGTCTACTCAGTGTGTCGAGGCCATTTGGCACTTCCAAT 183
DB 121 ATTGCTGATGTCGAAGATGAATTAGTCTACTCAGTGTGTCGAGGCCATTTGGCACTTCCAAT 180

QY 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTCTTAAATAATGCCGTGGACAAC 243
DB 181 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTCTTAAATAATGCCGTGGACAAC 240

QY 244 ACAGTTTCAACCTATGAAACATGGACATTATGTTCCAGCAATGCAGGAATTTCTGATCCC 303
DB 241 ACAGTTTCAACCTATGAAACATGGACATTATGTTCCAGCAATGCAGGAATTTCTGATCCC 300

QY 304 AACAGGCCCGCATATAGACACAGAAAGACACTTTGAAACGGGTTCTCAGTGTAAAT 363
DB 301 AACAGGCCCGCATATAGACACAGAAAGACACTTTGAAACGGGTTCTCAGTGTAAAT 360

QY 364 GTACCGGAGTTTTCCTATGTCATGAGCAGCAGCAGCTGTTATGATTCAGCACCCAGT 423
DB 361 GTACCGGAGTTTTCCTATGTCATGAGCAGCAGCAGCTGTTATGATTCAGCACCCAGT 420

QY 424 GGCAACATAATTTCCACTGCTAGTCTTAAGCTCAACTATGGTGGTGGTCTTCCACATGCC 483
DB 421 GGCAACATAATTTCCACTGCTAGTCTTAAGCTCAACTATGGTGGTGGTCTTCCACATGCC 480

QY 484 TATTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGA 543
DB 481 TATTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGA 540

QY 544 CAATTTGGCATTAGGTTAATCTTGTCTCTCTCTCCCTTCGGGCTTCTACGCTTTAGGCAAG 603
DB 541 CAATTTGGCATTAGGTTAATCTTGTCTCTCTCTCCCTTCGGGCTTCTACGCTTTAGGCAAG 600

QY 604 AAATTTTCAGGGATTAATAATGAAGAGAAATTTGAGAATGTAATAAATTTTCGGGGAAT 663
DB 601 AAATTTTCAGGGATTAATAATGAAGAGAAATTTGAGAATGTAATAAATTTTCGGGGAAT 660

QY 664 TTGAAGGTCCAAATTTAATGTGAGGATGTTGCCAATGCAGCTCTTATCTCGGCTAGT 723
DB 661 TTGAAGGTCCAAATTTAATGTGAGGATGTTGCCAATGCAGCTCTTATCTCGGCTAGT 720

QY 724 GATGAGGCAAAATACGTGAGTGACACAAATCTGTTCAATCATGAGGGTTACGCGTCTGC 783
DB 721 GATGAGGCAAAATACGTGAGTGACACAAATCTGTTCAATCATGAGGGTTACGCGTCTGC 780

QY 784 AATTCTGTATCAAGTGTTCCTCAATATCCAGATTCT 819
DB 781 AATTCTGTATCAAGTGTTCCTCAATATCCAGATTCT 816

RESULT 3
AAZ45731
ID AAZ45731 standard; cDNA; 831 BP.
XX AAZ45731;
AC AAZ45731;
XX
DT 06-APR-2000 (first entry)
XX
DE Secoisolariciresinol dehydrogenase cDNA clone SMDEHY321.
XX
KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;
KW enterolactone; enterodiol; nutraceutical; dietary supplement;
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
XX
OS Forsythia intermedia.
XX
XX Key Location/Qualifiers
FH 1.831
FT CDS
FT /*tag= a
FT /product= "secoisolariciresinol dehydrogenase"
FT /note= "no termination codon given"
XX
PN WO955846-A1.
XX
PD 04-NOV-1999.
XX
PF 23-APR-1999; 99WO-US08975.
XX
PR 24-APR-1998; 98US-0082977.
XX
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
PI Xia Z, Costa MA, Davin LB, Lewis NG;
XX
DR WPI; 2000-126356/11.
DR P-PSDB; AAY54413.
XX
PT New nucleic acid molecule encoding an enzyme involved in lignan
XX biosynthetic pathway, useful for producing large amounts of lignans
XX
PS Claim 9; Page 48-49; 66pp; English.
XX
CC The present sequence encodes a secoisolariciresinol dehydrogenase
CC protein. The enzyme is involved in the lignan biosynthetic pathway.
CC The secoisolariciresinol dehydrogenase proteins have a molecular
CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
CC nucleic acids are used for the recombinant expression of the enzymes.
CC It is also used to obtain expression or enhanced expression of
CC secoisolariciresinol dehydrogenase in plants or to alter lignan
CC biosynthesis. The enzyme is used for production of the pharmacologically
CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase
CC proteins and nucleic acids can be utilized to: elevate or otherwise
CC alter the levels of health-protecting lignans, including phytoestrogens
CC such as enterolactone and enterodiol, in plant species, including
CC vegetables, grains and fruits and to food items incorporating material
CC derived from such genetically altered plants; genetically alter plant
CC species to provide an abundant, natural supply of lignans useful for
CC a variety of purposes, for example as nutraceuticals and dietary
CC supplements; to genetically alter living organisms to produce an
CC abundant supply of optically pure lignans having desirable biological
CC properties, for example (-)-trachelogenin which possesses antiviral
CC properties, and (-)-podophyllotoxin.
XX
SQ Sequence 831 BP; 244 A; 169 C; 192 G; 226 T; 0 other;

Query Match 98.5%; Score 806.4; DB 21; Length 831;
Best Local Similarity 99.3%; Pred. No. 1.9e-226;
Matches 810; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 CAGCTTCGAAGTTCGCAAGAGGCTAGAGGAAAAAGTTGCCCTTATACAGAGGGA 63
DB 16 CAGGTTCTAACTGCAATCGCAAGAGGCTAGAGGAAAAAGTTGCCCTTATACAGAGGGA 75

QY 64 GCACGTGAATGGAGAAACACAGCAAACTCTCTCCACATGGAGCCAAAGTTGCC 123
DB 76 GCACGTGAATGGAGAAACACAGCAAACTCTCTCCACATGGAGCCAAAGTTGCC 135

QY 124 ATTGCTGATGCCAAGATGAATAGGTCACTAGTTGTGCGAGCCCATTTGGCACTTCCAAT 183
DB 136 ATTGCTGATGCCAAGATGAATAGGTCACTAGTTGTGCGAGCCCATTTGGCACTTCCAAT 195

QY 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAAC 243
DB 196 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAAC 255

QY 244 ACAGTTTCAACCTATGGAACACTGGACATTTATGTTTCAGCAATGCAGGAATTTCTGATCCC 303
DB 256 ACAGTTTCAACCTATGGAACACTGGACATTTATGTTTCAGCAATGCAGGAATTTCTGATCCC 315

QY 304 AACAGCCGCCCATCATAGACAACGAAAAAGCAGACTTTTGAACGGTCTCTCAGTGTAAAT 363
DB 316 AACAGCCGCCCATCATAGACAACGAAAAAGCAGACTTTTGAACGGTCTCTCAGTGTAAAT 375

QY 364 GTAACCGGAGTTTCCCTATGCATGAAGCAGCAGCAGCTGTTATGATTCACGACGCAAT 423
DB 376 GTAACCGGAGTTTCCCTATGCATGAAGCAGCAGCAGCTGTTATGATTCACGACGCAAT 435

QY 424 GGCACATAAATTCACATGCTAGTTTAAGCTCAACATATGGTGGTGGTTCACATGCC 483
DB 436 GGCACATAAATTCACATGCTAGTTTAAGCTCAACATATGGTGGTGGTTCACATGCC 495

QY 484 TATGCTGTTCAAGCATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGTCGGA 543
DB 496 TATGCTGTTCAAGCATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGTCGGA 555

QY 544 CAATTTGTCATAGGGTAAATGTTGTGTCCTTCCTCGGCTTCCCTACGGCTTTAGGCAAG 603
DB 556 CAATTTGTCATAGGGTAAATGTTGTGTCCTTCCTCGGCTTCCCTACGGCTTTAGGCAAG 615

QY 604 AAATTTTCAGGGATTAATAATGAAGAGAAATTTGAGAATGTATAAATCTTTCGCGGAAAT 663
DB 616 AAATTTTCAGGGATTAATAATGAAGAGAAATTTGAGAATGTATAAATCTTTCGCGGAAAT 675

QY 664 TTGAAAGTCCAAAATTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGGCTAGT 723
DB 676 CTGAAAGTCCAAAATTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGGCTAGT 735

QY 724 GATGAGGCAAAATACGTGAGTGAGCACAATCTGTTTCATTGATGGAGGTTTCAGGCTCTGC 783
DB 736 GATGAGGCAAAATACGTGAGTGAGCACAATCTGTTTCATTGATGGAGGTTTCAGGCTCTGC 795

QY 784 AATTCGTAAATCAAGTGTCCCAATATCCAGATTCT 819
DB 796 AATTCGTAAATCAAGTGTCCCAATATCCAGATTCT 831

RESULT 4
AAZ45734
ID AAZ45734 standard; cDNA; 828 BP.
XX
AC AAZ45734;
XX
XX
DT 06-APR-2000 (first entry)
XX Secoisolariciresinol dehydrogenase cDNA clone SMDEHY631.
XX Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;

KW enterolactone; enterodiol; neutriceutical; dietary supplement;
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
XX Forsythia intermedia.
OS
FH Key Location/Qualifiers
FT 1..828
FT /*tag= a
FT /product= "secoisolariciresinol dehydrogenase"
FT /note= "no termination codon given"
XX
PN WO9955846-A1.
XX
XX 04-NOV-1999.
PD
XX 23-APR-1999; 99WO-US08975.
PF
XX 24-APR-1998; 98US-0082977.
PR
XX (UNIW) UNIV WASHINGTON STATE RES FOUND.
PA
XX Xia Z, Costa MA, Davlin LB, Lewis NG;
PI
XX WPI: 2000-126356/11.
DR P-PSDB; AAY54416.
XX
XX New nucleic acid molecule encoding an enzyme involved in lignan
PT biosynthetic pathway, useful for producing large amounts of lignans
XX
XX Claim 9: Page 55-56; 66pp; English.
XX
CC The present sequence encodes a secoisolariciresinol dehydrogenase
CC protein. The enzyme is involved in the lignan biosynthetic pathway.
CC The secoisolariciresinol dehydrogenase proteins have a molecular
CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
CC nucleic acids are used for the recombinant expression of the enzymes.
CC It is also used to obtain expression or enhanced expression of
CC secoisolariciresinol dehydrogenase in plants or to alter lignan
CC biosynthesis. The enzyme is used for production of the pharmacologically
CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase
CC proteins and nucleic acids can be utilized to: elevate or otherwise
CC alter the levels of health-protecting lignans, including phytoestrogens
CC such as enterolactone and enterodiol, in plant species, including
CC vegetables, grains and fruits and to food items incorporating material
CC derived from such genetically altered plants; genetically alter plant
CC species to provide an abundant, natural supply of lignans useful for
CC a variety of purposes, for example as nutraceuticals and dietary
CC supplements; to genetically alter living organisms to produce an
CC abundant supply of optically pure lignans having desirable biological
CC properties, for example (-)-trachelogenin which possesses antiviral
CC properties, and (-)-podophyllotoxin.
XX
SQ Sequence 828 BP; 235 A; 172 C; 196 G; 225 T; 0 other;

Query Match 89.9%; Score 736; DB 21; Length 828;
Best Local Similarity 93.9%; Pred. No. 8.8e-206;
Matches 766; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 4 CAGCTTCGAAGTTCGCAAGAGGCTAGAGGAAAAAGTTGCCCTTATACAGAGGGA 63
DB 13 CAGCTTCGAAGTTCGCAAGAGGCTAGAGGAAAAAGTTGCCCTTATACAGAGGGA 72

QY 64 GCACGTGAATTTGGAGAAACACAGCAAACTCTTCTCCCAACATGGAGCCAAAGTTGCC 123
DB 73 GCACGTGAATTTGGAGAAACACAGCAAACTCTTCTCCCAACATGGAGCCAAAGTTGCC 132

QY 124 ATTGCTGATGTCCCAAGATGAATAGGTCACTAGTTGTGAGGCCCATTTGGCACTTCCAAT 183
DB 133 ATTGCTGATGTCCCAAGATGAATAGGTCACTAGTTGTGAGGCCCATTTGGCACTTCCAAT 192

QY 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAAC 243
DB 184 TCCACCTACATCCACTGTGATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACAAC 243

Db 1505 CAATTTGGCAATTAGGTTAATCTTTTCTCTCTTCCTTCGGGCTTCCTACGCCCTTTAGCCAAG 1564
QY 604 AAATTTTCAGGGATTAAATAATGAAGAAGAAATTTGAGAATGTAATAAACTTTGCGGGAAAT 663
Db 1565 AAATTTACAGGCAATTGAATAATGATGAAGACTTTGGCGAATGGAATGAACGTTGCGGGAAAT 1624
QY 664 TTGAAGGTCACAAATTAATGTTTGGAGTGTGGCAATGACGCTCTTATCTGGCTAGT 723
Db 1625 CTGAAGGTACAAATTTAGGATTTAGGATGTTGCCAATGACGCTCTTTTCTGGCTAGT 1684
QY 724 GATGAGCAAAATACGTCAGTGCACACATCTGTCATTTGATGGAGGTTTCAGGTCCTGC 783
Db 1685 GATGAGCCAAATATGTCAGTGGCAAAATCTGTCATGATGGAGGTTTCAGGTCCTGC 1744
QY 784 AATCTGTAATCAAAAGTGTCCAAATATCCAGATCT 819
Db 1745 AATCTGCAATCAAAATGTTTCCAATATCCAGACTCT 1780

RESULT 6
AAZ45732
ID AAZ45732 standard; cDNA; 819 BP.
XX
AC AAZ45732;
XX
DT 06-APR-2000 (first entry)
XX
DE Secoisolariciresinol dehydrogenase cDNA clone SMDEHY431.
XX
KW Secoisolariciresinol dehydrogenase; lignan biosynthetic pathway;
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;
KW enterolactone; enterodiol; neutriceutical; dietary supplement;
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
XX
OS Forsythia intermedia.
XX
FH Key Location/Qualifiers
FT CDS 1..819
FT /*tag= a
FT /product= "secoisolariciresinol dehydrogenase"
FT /note= "no termination codon given"
XX
PN WO955846-A1.
XX
PD 04-NOV-1999.
XX
PF 23-APR-1999; 99WO-US08975.
XX
PR 24-APR-1998; 98US-0082977.
XX
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
PI Xia Z, Costa MA, Davin LB, Lewis NG;
XX
PI WPI: 2000-126356/11.
DR P-FSDB; AAY34414.
XX
PT New nucleic acid molecule encoding an enzyme involved in lignan
PT biosynthetic pathway, useful for producing large amounts of lignans -
XX
PS Claim 9; Page 50-51; 66pp; English.
XX
CC The present sequence encodes a secoisolariciresinol dehydrogenase
CC protein. The enzyme is involved in the lignan biosynthetic pathway.
CC The secoisolariciresinol dehydrogenase proteins have a molecular
CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
CC NAD or NADP as a cofactor. The secoisolariciresinol dehydrogenase
CC nucleic acids are used for the recombinant expression of the enzymes.
CC It is also used to obtain expression or enhanced expression of
CC secoisolariciresinol dehydrogenase in plants or to alter lignan
CC biosynthesis. The enzyme is used for production of the pharmacologically
CC active lignan, matairesinol. The secoisolariciresinol dehydrogenase
CC proteins and nucleic acids can be utilized to: elevate or otherwise

CC alter the levels of health-protecting lignans, including phytoestrogens
CC such as enterolactone and enterodiol, in plant species, including
CC vegetables, grains and fruits and to food items incorporating material
CC derived from such genetically altered plants; genetically alter plant
CC species to provide an abundant, natural supply of lignans useful for
CC a variety of purposes, for example as nutraceuticals and dietary
CC supplements; to genetically alter living organisms to produce an
CC abundant supply of optically pure lignans having desirable biological
CC properties, for example (-)-trachelogenin which possesses antiviral
CC properties, and (-)-podophyllotoxin.
XX
SQ Sequence 819 BP; 235 A; 171 C; 191 G; 220 T; 2 other;

Query Match 89.6%; Score 733.8; DB 21; Length 819;
Best Local Similarity 93.4%; Pred. NO. 3.9e-205;
Matches 765; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1 ATGCAGCTTCGAACCTGCAATTCGCAAGAGCTAGAAAGAAAAGTTGCCCTTATAACAGGA 60
Db 1 ATGCAGCTTCGAACCTGCAATTCGCAAGAGCTAGAAAGAAAAGTTGCCCTTATAACAGGA 60

QY 61 GGAGCCAGTGAATTTGGAGAAACACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTT 120
Db 61 GGAGCCAGTGGAGTTGGAGAAATCAGAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTT 120

QY 121 GCCATTGCTGATGTCCAAAGATGAATTAGTGCACAGTGTTCGAGGCCATTTGGCCCTTCC 180
Db 121 GCCATTGCTGATGTCCAAAGATGAATTAGTGCACAGTGTTCGAGGCCATTTGGCCCTTCC 180

QY 181 AATTCACCTACATCCACTGCTGATGTTTACTTAATGAAGACGGTGTAAAAATGCCGTGAC 240
Db 181 AATTCACCTACATCCACTGCGATGTTTACTTAATGAAGACGGTGTAAAAATGCCGTGAC 240

QY 241 AACAGGTTTCAACCTATGGAACACTGGACATTTATGTTTCAGCAATGCAGGAATTTCTGAT 300
Db 241 AACAGGTTTCAACCTATGGAACACTGGACATTTATGTTTCAGCAATGCAGGAATTTCTGAT 300

QY 301 CCCACAGGCCCCGCATCATAGACAAAGAAAAGCAGACTTTTGAACGGTTCCTCAGTGA 360
Db 301 CCCACAGGCCCCGGGTCTAGTAAAGCTCAACTATGGTGGTGGTTCCTCAGTGA 360

QY 361 AATGTACCGGAGTTTTCCTATGCATGAACGACGACGACGCTGTTATGATTTCCAGCAGC 420
Db 361 AATGTACCGGAGTTTTCCTATGTTTATGAACGACGACGACGCTGTTATGTTCCAGCAGC 420

QY 421 AGTGGCAACATAATTTCCACTGCTAGTTTAAAGCTCAACTATGGTGGTGGTTCCTCAGT 480
Db 421 AATGGCTGCATAATTTCCACTGCTAGTTTAAAGCTCAACTATGGTGGTGGTTCCTCAGT 480

QY 481 GCCTATTGTGTTCAAAAGCATGCTGTGTTAGCCCTTACTAGGAATCTGCAGTCGAGCTC 540
Db 481 GCTTATTGTGGTCAAAACATGCTGTATTAGGCCCTTACTAGGAATCTGCAGTCGAGCTC 540

QY 541 GGACAATTTGGCATTAGGGTTAATGTTTGTCTCTTCCTTCGGGCTTCCTACGGCTTTAGGC 600
Db 541 GGACAATTTGGCATTAGGGTTAATGTTTGTCTCTTCCTTCGGGCTTCCTACGGCTTCAGC 600

QY 601 AAGAAATTTTCAGGGATTAAAAATGAAGAAGAAATTTGAGAATGTAATAAACTTTCCGGGA 660
Db 601 AAGAAATTTTCAGGGATTAAAAATGATGTAGACTTTTGCAGAAATGCAATAGAACATCGGGA 660

QY 661 AATTGAAAGTCCAAAATTTAATGTTTGGAGTGTTCGCAATGCCAGCTCTTTTATCTGGCT 720
Db 661 AATCTGAAAGGTACAAAATTTGAGGATTTGAGGATTTGAGGATTTGAGGATTTGAGGATTT 720

QY 721 AGTATGAGGCAAAATACGTCAGTGGACACAACTCTCTTCTTATGATGGAGGTTTCAGCTC 780
Db 721 AGTATGAGGCAAAATATGTCAGTGGACACAAATCTCTTCTTATGATGGAGGTTTCAGCTC 780

QY 781 TGCATCTCTTAATCAAGTGTTCCTCAATATCCAGATCT 819
Db 781 TGCATCTCTTAATCAAGTGTTCCTCAATATCCAGATCT 819

RESULT 7
AAZ45733
ID AAZ45733 standard; cDNA; 831 BP.
XX
AC AAZ45733;
XX
DT 06-APR-2000 (first entry)
XX
DE Secoisolaricresinol dehydrogenase cDNA clone SMDEHY511.
XX
KW Secoisolaricresinol dehydrogenase; lignan biosynthetic pathway;
KW lignan; matairesinol; health-protecting lignan; phytoestrogen;
KW enterolactone; enterodiol; neutriceutical; dietary supplement;
KW (-)-trachelogenin; antiviral; (-)-podophyllotoxin; ss.
XX
OS Forsythia intermedia.
XX
FH Key Location/Qualifiers
CDS 1..831
FT /*tag= a
FT /product= "secoisolaricresinol dehydrogenase"
FT /note= "no termination codon given"
XX
XX WO955846-A1.
XX
PD 04-NOV-1999.
XX
XX 23-APR-1999; 99WO-US08975.
XX
XX 24-APR-1998; 98US-0082977.
XX
XX (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
XX Xia Z, Costa MA, Davin LB, Lewis NG;
XX
XX WPI; 2000-126356/11.
DR P-PSDB; AAY54415.
XX
PT New nucleic acid molecule encoding an enzyme involved in lignan
PT biosynthetic pathway, useful for producing large amounts of lignans -
XX
XX Claim 9; Page 52-54; 66pp; English.
XX
CC The present sequence encodes a secoisolaricresinol dehydrogenase
CC protein. The enzyme is involved in the lignan biosynthetic pathway.
CC The secoisolaricresinol dehydrogenase proteins have a molecular
CC weight of 27-31 kDa, an isoelectric point of 5.9-6.85, and require
CC NAD or NADP as a cofactor. The secoisolaricresinol dehydrogenase
CC nucleic acids are used for the recombinant expression of the enzyme.
CC It is also used to obtain expression or enhanced expression of
CC secoisolaricresinol dehydrogenase in plants or to alter lignan
CC biosynthesis. The enzyme is used for production of the pharmacologically
CC active lignan, matairesinol. The secoisolaricresinol dehydrogenase
CC proteins and nucleic acids can be utilized to: elevate or otherwise
CC alter the levels of health-protecting lignans, including phytoestrogens
CC such as enterolactone and enterodiol, in plant species, including
CC vegetables, grains and fruits and to food items incorporating material
CC derived from such genetically altered plants; genetically alter plant
CC species to provide an abundant, natural supply of lignans useful for
CC a variety of purposes, for example as neutriceuticals and dietary
CC supplements; to genetically alter living organisms to produce an
CC abundant supply of optically pure lignans having desirable biological
CC properties, for example (-)-trachelogenin which possesses antiviral
CC properties, and (-)-podophyllotoxin.
XX
SQ Sequence 831 BP; 244 A; 181 C; 195 G; 211 T; 0 other;

Query Match 86.9%; Score 712; DB 21; Length 831;
Best Local Similarity 92.0%; Pred. No. 9 8e-199;
Matches 751; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 4 CAGCTTCGAACGTCATTCGCAAGAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGA 63

Db 16 CAGGTTCTAACTGCAATCAAGAAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGA 75
QY 64 GCCAGTGGAAATGGAGAAACACAGCAAAACACTCTCTCCCAACATCGAGCCAAAGTTGCC 123
Db 76 GCCAGTGGAAATGGAGAAATTCACAGCAAAACACTCTCTCCCAACATCGAGCCAAAGTTGCC 135
QY 124 ATTGCTGATGTCCAAGATGAATTAGTCACTCAGTTGTGCGAGGCCATTGGCACTTCCAAT 183
Db 136 ATTGCTGATGTCCAAGATGAATTAGTCACTCAGTTGTGCGAGGCCATTGGCACTTCCAAT 195
QY 184 TCACCTACATCCACTGTGATGTTTACTAATGAAGAGCGGTGTTAAAAATCCCGTGGACAAC 243
Db 196 TCCATCTACATCCACTGCGATGTTACCAATGAAGACGATGTTAAAAATCCCGTGGACAAC 255
QY 244 ACAGTTTCAACCTATGGAACACTGGACATTATCTTCAGCAATGCGAGAAATTTCTGTATCC 303
Db 256 ACAGTTTCAACCTATGGAACACTGGACATTATCTTCACAAATGCGAGAAATTTCTGTATCC 315
QY 304 AACAGCCCCGCATCATAGACAACGAAAAAGCAGACTTTGNAAGCGCTTCTCAGTGTAAAT 363
Db 316 AACAGCCCCGCATCGTAGACAACGAAAAAGCAGACTTTGNAAGCGCTTCTCAGCGTAAAT 375
QY 364 GTAACCGGAGTTTCTCTATGCAATGAAGCAGCAGCAGCGTTATGATGTTCCAGCAGCCAGT 423
Db 376 GTAACCGGTTTCTCTATGCAATGAAGCAGCAGCAGCGTTATGATGTTCCAGCAGCCAGT 435
QY 424 GGCAACATAATTTCCACTGCTAGTTTAAAGCTCAACTATATGGTGGTGGTTTTCACATGCC 483
Db 436 GGCAGCATAAATTTCCACTGCTAGCTAAAGCTCAACAAATGGTGGTGGTGGTGGTGGTGGT 495
QY 484 TATTTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGCGAGCTCGAGCTCGGA 543
Db 496 TATTTGGTTCAAAGCATGCTGTGTAGCCCTTACTAGGAATCTGCGAGCTCGGA 555
QY 544 CAATTTGGCATTAGGTTAATTTGTTGTTCTCTTTCGGGCTTCTTACGGCTTTTAGGCAAG 603
Db 556 CAATTTGGCATTAGGTTAATTTGTTGTTCTCTTTCGGGCTTCTTACGGCTTTTAGGCAAG 615
QY 604 AAATTTTCAGGGATTAATAAATGAAGAAGAAATTTGACAATGTAATAAATTTTCGGGAAAT 663
Db 616 AAATTTTCAGGGCTTGAATAATGACGAAGATTTGGAGAATGCAATGAGCCCTTATGGGAAAT 675
QY 664 TTGAAAGTCCAAAATTTAATGTTGAGGATGTTGCCAATGCGAGCTCTTTATCTGGCTAGT 723
Db 676 CTGAAAGGTACAAATTTGAAGCTGAGGAGCTGCGCAATGCGAGCTCTTTATCTGGCAAGT 735
QY 724 GATGAGGCAAAATACGTGAGTGAGACACAATCTGTTCAATGATGGAGGGTTACGCGTCTGC 783
Db 736 GATGAGGCAAAATATGTGAGTGAGACACAATCTGTTCAATGATGGAGGGTTACGCGTCTAC 795
QY 784 AATTTCTGTAATCAAAAGTGTTCCAATATCCAGATTCT 819
Db 796 AATTTCTGCAATCAAAATGTTCCAATATCCAGACT 831

RESULT 8
AAC40811
ID AAC40811 standard; DNA; 1120 BP.
XX
AC AAC40811;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 29624.
DE Arabidopsis thaliana
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
PN EP1033405-A2.

XX	06-SEP-2000.	PR	02-JUL-1999;	99US-0142055.
PD		PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
PF	25-FEB-2000;	PR	09-JUL-1999;	99US-0142820.
XX	2000EP-0301439.	PR	12-JUL-1999;	99US-0142877.
XX		PR	13-JUL-1999;	99US-0143342.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	04-MAY-1999;	PR	21-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	07-MAY-1999;	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	PR	03-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	PR	10-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	PR	12-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	PR	13-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	24-JUN-1999;	PR	01-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	PR	07-SEP-1999;	99US-0152363.
PR	29-JUN-1999;	PR	10-SEP-1999;	99US-0153070.
PR	30-JUN-1999;	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	PR	15-SEP-1999;	99US-0154018.
PR		PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	PR	20-SEP-1999;	99US-0154779.

QY	7	CTTCGAATCGATTTCGCAAGAAGCTAGAGAAAAGTTGCCCTTATACAGGAGGAGCC	66	PN	EP1033405-A2.
Db	47	CTTCATCACCATAAGCAAAAGCTAGAAGGTAAAGTAGCAGCTAATAACTGGTGGAGCT	106	XX	06-SEP-2000.
QY	67	AGTGAATTGGAGAAACACAGCAAAACTCTTCTCCACATGGAGCCAAAGTTGCCATT	126	XX	25-FEB-2000; 2000EP-0301439.
Db	107	AGCGCATAGAGCGCGCCACAGCTAGCTTTTGTTCACATGGTGGAAAGTTACAATT	166	XX	99US-0121825.
QY	127	GCTGATGCCAAGATGAATTAGTCTACTAGTTGTCGAGGCCATTGGCAGCTTCCAATTCC	186	PR	99US-0123360.
Db	167	CGAGACATCAAGACACACCTTGGAACTCTTTAGTACAAGAAATTTGGCAACGACATACA	226	PR	99US-0123548.
QY	187	ACCTACATCCACTGTGATGTACTTAATGAAGACGGTGTAAATAATGCCGTGGACAAACA	246	PR	99US-0125788.
Db	227	ATCTTTATCCATTCGAATGTGTGTGTAATCAGACGTTCAAAATGTTGTAGATGCAACA	286	PR	99US-0126264.
QY	247	GTTTCAACCTATGGAACACTGGACATTATGTTTCAGCAATGTCAGGAAATTTCTGATCCCAAC	306	PR	99US-0126785.
Db	287	ATTGCCAAATTTGGTAAGCTGGACATAATGTTTCAGTAGCGTGGTATAGGAGTAAGTCA	346	PR	99US-0127462.
QY	307	AGGCCCGCATCATAGACAACGAAAAGCAGACTTTTGAACGCGTTCTCAGTGTAAATGTA	366	PR	99US-0128234.
Db	347	ATTTCCAGCATCTTAGATGTCGTACGAATA---ATTAAGACAGCTGTCGATGTAACATT	403	PR	99US-0128714.
QY	367	ACCGGAGTTTCCCTATGCATGAAGCAGCAGCAGCGTTTATGATTCCA-GCACGCGATGG	425	PR	99US-0129845.
Db	404	GTTGCGCCTCTTTTCGCGCTAAGCATGCTGCTAGAGTAATGATTCATTCATCAAGAAAGG	463	PR	99US-0130077.
QY	426	CAACATAATTTCCACTGCTAGTTTAAGCTCAACTATGGGTGGTCTTTCACATGCCCTA	485	PR	99US-0130449.
Db	464	TTCCATATCTTCACACAGATTAATGTTC-----CCTCAGCGCAT	503	PR	99US-0130510.
QY	486	TTGTGTTTCAAAGCA-TGCTGTGTTAGCCCTTACTAGGAATCTGCACTGCGAGTCTGGAC	544	PR	99US-0130891.
Db	504	TTCTGTCATCAAAAAGGTGCGAGTTTGGGATTTTCCCAAGAACATTTGGGTGCAATTAGGGA	563	PR	99US-0131449.
QY	545	AATTTGGCATTAGGTTAAATGTTGTTCTCTCTTCGCGCTTCCCTACGCTTTAGGCAAGA	604	PR	99US-0132048.
Db	564	AGTACGGAATAAAATGTAAGTGTGTTCTCCTCAATTACATTAGCACACCACCTTGATTGA	623	PR	99US-0132485.
QY	605	AATTTTCAGGATTAATAATGAAGAAGATTTGAGAAATGTAATAAATTTGCGGGAATTT	664	PR	99US-0132486.
Db	624	ATGCGCTTGAATAAGCTGAGAGAGAGATAGCGGAAAATGGTTTCAGCAGGAGGAAAT	683	PR	99US-0132487.
QY	665	TGAAAGTCCAAAATTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGCTAGTG	724	PR	99US-0132863.
Db	684	TGAAAGGAGCTTTACTAGATGAAGAGGAGTCCGAAAGCAGTCTATCT-TAGCAAGTG	742	PR	99US-0134218.
QY	725	ATGAGGCAAAATAGTGCAGTGACACAATCTGTTCAATTGATGAGGGTTCAGCGTCTGCA	784	PR	99US-0134219.
Db	743	ATGATTCTAAATATGCGTGGTATGAATCTGGTTATTGATGGTGGTTTTAGTACCACAA	802	PR	99US-0134221.
QY	785	AT 786		PR	99US-0134370.
Db	803	AT 804		PR	99US-0134768.
RESULT 10				PR	99US-0134941.
AAC38191				PR	99US-0135124.
ID	AAC38191	standard; DNA; 1136 BP.		PR	99US-0135353.
XX				PR	99US-0135629.
AC	AAC38191;			PR	99US-0136021.
XX				PR	99US-0136392.
DT	17-OCT-2000	(first entry)		PR	99US-0136782.
XX				PR	99US-0137222.
DE	Arabidopsis thaliana	DNA fragment SEQ ID NO: 20099.		PR	99US-0137528.
KW	Hybridisation assay; genetic mapping; gene expression control;			PR	99US-0137724.
KW	protein identification; signal transduction pathway;			PR	99US-0138094.
XX	metabolic pathway; promoter; termination sequence; ss.			PR	99US-0138540.
OS	Arabidopsis thaliana.			PR	99US-0138847.
XX				PR	99US-0139119.
XX				PR	99US-0139452.
XX				PR	99US-0139453.
XX				PR	99US-0139454.
XX				PR	99US-0139455.
XX				PR	99US-0139456.
XX				PR	99US-0139457.
XX				PR	99US-0139458.
XX				PR	99US-0139459.
XX				PR	99US-0139460.
XX				PR	99US-0139461.
XX				PR	99US-0139462.
XX				PR	99US-0139463.
XX				PR	99US-0139750.
XX				PR	99US-0139763.
XX				PR	99US-0139817.
XX				PR	99US-0139899.
XX				PR	99US-0140353.
XX				PR	99US-0140354.
XX				PR	99US-0140695.
XX				PR	99US-0140823.
XX				PR	99US-0140991.
XX				PR	99US-0141287.
XX				PR	99US-0141842.

PR	01-JUL-1999;	9905-0142154
PR	02-JUL-1999;	9905-0142055
PR	06-JUL-1999;	9905-0142390
PR	08-JUL-1999;	9905-0142803
PR	09-JUL-1999;	9905-0142920
PR	12-JUL-1999;	9905-0142977
PR	13-JUL-1999;	9905-0143542
PR	14-JUL-1999;	9905-0143624
PR	15-JUL-1999;	9905-0144005
PR	16-JUL-1999;	9905-0144085
PR	16-JUL-1999;	9905-0144086
PR	19-JUL-1999;	9905-0144325
PR	19-JUL-1999;	9905-0144331
PR	19-JUL-1999;	9905-0144332
PR	19-JUL-1999;	9905-0144333
PR	19-JUL-1999;	9905-0144334
PR	19-JUL-1999;	9905-0144335
PR	20-JUL-1999;	9905-0144352
PR	20-JUL-1999;	9905-0144632
PR	20-JUL-1999;	9905-0144884
PR	21-JUL-1999;	9905-0144814
PR	21-JUL-1999;	9905-0145086
PR	21-JUL-1999;	9905-0145088
PR	22-JUL-1999;	9905-0145085
PR	22-JUL-1999;	9905-0145087
PR	22-JUL-1999;	9905-0145089
PR	22-JUL-1999;	9905-0145192
PR	23-JUL-1999;	9905-0145145
PR	23-JUL-1999;	9905-0145218
PR	23-JUL-1999;	9905-0145224
PR	26-JUL-1999;	9905-0145276
PR	27-JUL-1999;	9905-0145913
PR	27-JUL-1999;	9905-0145918
PR	27-JUL-1999;	9905-0145919
PR	28-JUL-1999;	9905-0145951
PR	02-AUG-1999;	9905-0146386
PR	02-AUG-1999;	9905-0146388
PR	03-AUG-1999;	9905-0146389
PR	03-AUG-1999;	9905-0147038
PR	04-AUG-1999;	9905-0147204
PR	04-AUG-1999;	9905-0147302
PR	05-AUG-1999;	9905-0147192
PR	06-AUG-1999;	9905-0147260
PR	06-AUG-1999;	9905-0147303
PR	06-AUG-1999;	9905-0147416
PR	09-AUG-1999;	9905-0147493
PR	09-AUG-1999;	9905-0147935
PR	10-AUG-1999;	9905-0148171
PR	11-AUG-1999;	9905-0148319
PR	12-AUG-1999;	9905-0148341
PR	13-AUG-1999;	9905-0148565
PR	13-AUG-1999;	9905-0148684
PR	16-AUG-1999;	9905-0149368
PR	17-AUG-1999;	9905-0149175
PR	18-AUG-1999;	9905-0149426
PR	20-AUG-1999;	9905-0149722
PR	20-AUG-1999;	9905-0149723
PR	20-AUG-1999;	9905-0149929
PR	23-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150884
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	27-AUG-1999;	9905-0151080
PR	30-AUG-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151348
PR	01-SEP-1999;	9905-0151930
PR	07-SEP-1999;	9905-0152363
PR	10-SEP-1999;	9905-0153070
PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039

QY	443	CPAGTTTAAGCTCAACATATAGGGTGGTTCCTCAACATCCCTATTGTTGTTCAAAAGCATG	502	PR	14-MAY-1999;	99US-0134370.
				PR	18-MAY-1999;	99US-0134768.
Db	711	GGAGTGTACGGGATGATGGGGTTTACGCCAACATACTTACACGCTCTCAAAATCCG	770	PR	19-MAY-1999;	99US-0134941.
				PR	20-MAY-1999;	99US-0135124.
QY	503	CTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGGCATTAGGGTTA	562	PR	21-MAY-1999;	99US-0135353.
				PR	24-MAY-1999;	99US-0135629.
Db	771	CTGTTATCGGAATTGTAAAGATCAACAGCTTCAGAACTATGCAAGCACAGGATCCGGTCA	830	PR	25-MAY-1999;	99US-0136021.
				PR	27-MAY-1999;	99US-0136392.
QY	563	ATTGTTTCTCCTTTTCGGCTTCTACGGCTTTAGGCAAGAAAT-----TTT	610	PR	28-MAY-1999;	99US-0136782.
				PR	01-JUN-1999;	99US-0137222.
Db	831	ACTGCAATTCCTCTTTTGGCATCACACATCATTCGTGATGATGATGCCACAGATTT	890	PR	03-JUN-1999;	99US-0137528.
				PR	04-JUN-1999;	99US-0137502.
QY	611	CAGGGAATTAATAATCAAGACAAATTTGAGNATGTATAAATTTGCGGAAATTTGAAG	670	PR	07-JUN-1999;	99US-0137724.
				PR	08-JUN-1999;	99US-0138094.
Db	891	ACCCCGTGTGTGATGATCTCAGGCTGATCCAGATGTCAGAGTACAGGTGTTAAATG	950	PR	10-JUN-1999;	99US-0138540.
				PR	14-JUN-1999;	99US-0138847.
QY	671	GTCCAAAATTTAATGTTGAGATGTTGCCAATGCAGCTTTATCTGCGTAGTGATGAGG	730	PR	16-JUN-1999;	99US-0139119.
				PR	17-JUN-1999;	99US-0139452.
Db	951	GAGAGTTTGTGAACCAACCGATGTAGCTAATGCACGCGTGTATCTCGCTTCCGATGATT	1010	PR	18-JUN-1999;	99US-0139453.
				PR	17-JUN-1999;	99US-0139492.
QY	731	CAAAATAGCTGAGTGGACACAATCTGTTCAATGATGAGGGTTCA	775	PR	18-JUN-1999;	99US-0139454.
				PR	18-JUN-1999;	99US-0139455.
Db	1011	CAAAATATGTAATGGGCATAATCTGTTGATGATGAGGATTTCA	1055	PR	18-JUN-1999;	99US-0139456.
				PR	18-JUN-1999;	99US-0139457.
				PR	18-JUN-1999;	99US-0139458.
				PR	18-JUN-1999;	99US-0139459.
				PR	18-JUN-1999;	99US-0139460.
				PR	18-JUN-1999;	99US-0139461.
				PR	18-JUN-1999;	99US-0139462.
				PR	18-JUN-1999;	99US-0139463.
				PR	18-JUN-1999;	99US-0139750.
				PR	18-JUN-1999;	99US-0139763.
				PR	21-JUN-1999;	99US-0139817.
				PR	22-JUN-1999;	99US-0139899.
				PR	23-JUN-1999;	99US-0140353.
				PR	23-JUN-1999;	99US-0140354.
				PR	24-JUN-1999;	99US-0140695.
				PR	28-JUN-1999;	99US-0140823.
				PR	29-JUN-1999;	99US-0140991.
				PR	30-JUN-1999;	99US-0141287.
				PR	01-JUL-1999;	99US-0141842.
				PR	01-JUL-1999;	99US-0142154.
				PR	02-JUL-1999;	99US-0142055.
				PR	06-JUL-1999;	99US-0142390.
				PR	08-JUL-1999;	99US-0142803.
				PR	09-JUL-1999;	99US-0142920.
				PR	12-JUL-1999;	99US-0142977.
				PR	13-JUL-1999;	99US-0143542.
				PR	14-JUL-1999;	99US-0143624.
				PR	15-JUL-1999;	99US-0144005.
				PR	16-JUL-1999;	99US-0144085.
				PR	16-JUL-1999;	99US-0144086.
				PR	19-JUL-1999;	99US-0144325.
				PR	19-JUL-1999;	99US-0144331.
				PR	19-JUL-1999;	99US-0144332.
				PR	19-JUL-1999;	99US-0144333.
				PR	19-JUL-1999;	99US-0144334.
				PR	19-JUL-1999;	99US-0144335.
				PR	20-JUL-1999;	99US-0144352.
				PR	20-JUL-1999;	99US-0144632.
				PR	20-JUL-1999;	99US-0144884.
				PR	21-JUL-1999;	99US-0144814.
				PR	21-JUL-1999;	99US-0145086.
				PR	21-JUL-1999;	99US-0145088.
				PR	22-JUL-1999;	99US-0145085.
				PR	22-JUL-1999;	99US-0145087.
				PR	22-JUL-1999;	99US-0145089.
				PR	22-JUL-1999;	99US-0145192.
				PR	23-JUL-1999;	99US-0145145.
				PR	23-JUL-1999;	99US-0145218.
				PR	23-JUL-1999;	99US-0145224.
				PR	26-JUL-1999;	99US-0145276.
				PR	27-JUL-1999;	99US-0145913.

RESULT 11
AAC43068
ID AAC43068 standard; DNA: 852 BP.
XX
AC AAC43068;
XX
DT
XX
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 37909.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 47157.
DE Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140951.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR	27-AUG-1999;	99US-0151080.	Query Match	22.4%;	Score 183.6;	DB 21;	Length 1138;
PR	30-AUG-1999;	99US-0151303.	Best Local Similarity	54.7%;	Pred. No. 1.7e-43;		
PR	31-AUG-1999;	99US-0151438.	Matches 416;	Conservative	0;	Mismatches 329;	Indels 15; Gaps 2;
PR	01-SEP-1999;	99US-0151930.					
PR	07-SEP-1999;	99US-0152363.					
PR	10-SEP-1999;	99US-0153070.					
PR	13-SEP-1999;	99US-0153758.					
PR	15-SEP-1999;	99US-0154018.					
PR	16-SEP-1999;	99US-0154039.					
PR	20-SEP-1999;	99US-0154779.					
PR	22-SEP-1999;	99US-0155139.					
PR	23-SEP-1999;	99US-0155486.					
PR	24-SEP-1999;	99US-0155659.					
PR	28-SEP-1999;	99US-0156458.					
PR	29-SEP-1999;	99US-0156596.					
PR	04-OCT-1999;	99US-0157117.					
PR	05-OCT-1999;	99US-0157753.					
PR	06-OCT-1999;	99US-0157865.					
PR	07-OCT-1999;	99US-0158029.					
PR	08-OCT-1999;	99US-0158232.					
PR	12-OCT-1999;	99US-0158369.					
PR	13-OCT-1999;	99US-0159293.					
PR	13-OCT-1999;	99US-0159294.					
PR	13-OCT-1999;	99US-0159295.					
PR	14-OCT-1999;	99US-0159329.					
PR	14-OCT-1999;	99US-0159330.					
PR	14-OCT-1999;	99US-0159331.					
PR	14-OCT-1999;	99US-0159637.					
PR	14-OCT-1999;	99US-0159638.					
PR	18-OCT-1999;	99US-0159584.					
PR	21-OCT-1999;	99US-0160741.					
PR	21-OCT-1999;	99US-0160767.					
PR	21-OCT-1999;	99US-0160768.					
PR	21-OCT-1999;	99US-0160770.					
PR	21-OCT-1999;	99US-0160814.					
PR	21-OCT-1999;	99US-0160815.					
PR	22-OCT-1999;	99US-0160980.					
PR	22-OCT-1999;	99US-0160981.					
PR	22-OCT-1999;	99US-0160989.					
PR	25-OCT-1999;	99US-0161404.					
PR	25-OCT-1999;	99US-0161405.					
PR	25-OCT-1999;	99US-0161406.					
PR	26-OCT-1999;	99US-0161359.					
PR	26-OCT-1999;	99US-0161360.					
PR	26-OCT-1999;	99US-0161361.					
PR	28-OCT-1999;	99US-0161920.					
PR	28-OCT-1999;	99US-0161992.					
PR	28-OCT-1999;	99US-0161993.					
PR	29-OCT-1999;	99US-0162142.					

QY	28	AGCTAGAGAAAGTTGCCCTTATAACAGAGAGCCAGCTGGAATTGGAGAAACACA	87
DB	303	AAAGTAAAGGTAAGTAGCACTCATCTAGGAGGAGCAAGTGGGATTGGCAAGCAACA	362
QY	88	GCAAACTCTTCCCAACATGAGGCAAGTGCCTTGCCTGATGCTCAAGATGAATTA	147
DB	363	GGCGAAAATTCATCATGCTAGTGGAGCAAGTTATCATTTGCCGATATCCAAACGAGATT	422
QY	148	GGTCACTAGTTTCGAGGCAATTCACATTCACATTCACATTCACATTCATGATGT	207
DB	423	GGCGGAGAACCCGAGCAAGTCCGG---TCCAGTTGTCTTCTCCATCCGATGTG	479
QY	208	ACTAATGAGACGGTGTAAAATGCGGTGGACACACAGATTTCAACCTATGGAACATG	267
DB	480	ACCAAGAAATCAGACATTCCTAACGCGAGTTGACTTCGCTGCTCGCTCCATCAAGCTC	539
QY	268	GACATTATGTTACGAAATCAGGAATTTCTGATCCCAACAGGCGCCGATCATAGACAC	327
DB	540	GACATTATGTAACAACATCTGTTATTCCTGCAAAACGCCCTCCTAGTATCGTTGATCTT	599

QY	328	GAAGAAGCAGACTTTGAACGCGTTCTCAGTGAATGTAAACCGAGTTTCTCTATCATG	387
DB	600	GATCTCAATGTTTCGACAAGGTAATCAACACAATGTCGCTGGAGTCATGCGAGAATC	659
QY	388	AAGCAGCAGCAGCGTGTATGATTCAGCAGCAGCAGTGGCAACATAATTTCCACTGCTAGT	447
DB	660	AAACATGCTGCTCGTGTGATGATCCCGCTAACTCTGGATCCATCATTTTGTGAGGAGCT	719
QY	448	TTAAGCTCAACTATGGTGGTGGTCTTTCACATGCTTATTTGGTTCAAGCATGCTGTG	507
DB	720	GTCACGGGATGATGGGCGGTTTAGCCCAACATCTACAGCGTCTCAAAAATCCCGTGT	779
QY	508	TTAGCCCTTACTAGGAATCTGCAGTCGAGCTCGGACAATTTGGCATTAGGGTTAATTTGT	567
DB	780	ATCGGAATTTGAAGTCAACACAGCTTCAGAACTATGCAACACAGGATCCGGGTCAACTGC	839
QY	568	TTGTCTCTTTTCGGGCTTCTACGGCTTTAGGCAAGAAAT-----TTTCAGGG	615
DB	840	ATTTCCTCTTTTCGGATCACACATCATTCGTGATGGATGAGATGCGACAGATTTACCCC	899
QY	616	ATTAAAAATGAAGAAGAAATTGAGATGTAATAAACTTTGCGGGAATTTGAAAGTCCA	675
DB	900	GGTGTGATGACTCAAGGCTGATCCAGATAGTCAGAGTACAGAGTGTAAATGGAGAG	959
QY	676	AAATTTAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGCTAGTGTATGATGAGGCAAAA	735
DB	960	CTTTGTGAACCAACCGATGCTAGCTAATGCGCGGTATCTCGCTCCCATGATTCAAAG	1019
QY	736	TACGTAGTGGACACAATCTGTTCAATGATGAGGGTTCA	775
DB	1020	TATGTAATGGGCATAATCTGCTGGTAGATGGAGGATTCA	1059

RESULT 13
AAC43159
ID AAC43159 standard; DNA; 774 BP.
XX
AC AAC43159;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 38243.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	11-MAY-1999;	99US-0132863.
PR	14-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137702.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	14-JUN-1999;	99US-0138847.
PR	16-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	17-JUN-1999;	99US-0139453.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	24-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	09-JUL-1999;	99US-0142803.
PR	12-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144684.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149920.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.

PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 18.3%; Score 150.2; DB 21; Length 774;			
Best Local Similarity 52.7%; Pred. No. 8.9e-34;			
Matches 401; Conservative 0; Mismatches 348; Indels 12; Gaps 3;			
QY	28	AGCGTAGAGGAAGCTGCCCTTATACAGGAGGAGCCAGTGGAAATTCGAGAAACCA 87	
Db	13	AGATTGGATGGCAAAATCGCAATAATAACAGCGGAGCTAGCGGGATTGGAGCCGAAGCG 72	
QY	88	GCAAAACTCTTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGTCACCAAGATGAATTA 147	
Db	73	GTTAGCTGTTTCAGGACCCAGCGAGCTAGGTGTCATCGTTGACTTTCAGAGAGACTT 132	
QY	148	GGTCACTAGTTTCGAGGCCATTTGCGACTTTCCAATTCACCTTACATCCACTGTGATGTT 207	
Db	133	GGTCAAAACGTTGCCGTTTCTGTCGGGAAGACAAAGCAAGTTTTTACCGTTTCGGATGTT 192	
QY	208	ACTAATGAACAGCGTCTTAAATCCGTCGACCAACACAGTTTCAACCTATGCAAAACTG 267	
Db	193	ACAAAGGAAAGGAGGTAGAAAACGCGTTAAGTTTACCGTCGAAAAGTACGGGAAGCTT 252	
QY	268	GACATTATGTTACGAATCAGGAATTTCTGATCCCAACAGGCCCGCATATAGACAAC 327	
Db	253	GACGTTCTCTTAGTAACCGCGCTTATGGAAC---AGCCGGGAAGCTTCTCGACTTG 309	
QY	328	GAAAAGCAGACTTTGAACGCGTTCTCAGTGTAAATGTAAACCGAGATTTTCTTATGCAATG 387	
Db	310	AATCTGGAACAGTTTGACCGCAACCATGGCGGTCAACGTTTCGTGGTGGCGTGGCTTATC 369	
QY	388	AAGCAGCAGCAGCTGTTATGATTCAGCACCA---GTGGCAACATAATTTCCACTGCT 444	
Db	370	AAACGCTGCACGACCCATGGTGGAGAAAGGCACGCGTGGGTCAATCGTATGTACGACC 429	
QY	445	AGTTTAAGCTCAACTATGGGTGGTGTCTTTCACATGCCTATTGTGTTCAAAGCATGCT 504	
Db	430	AGGTCGCGTCGAGATCGGTGTCTCAGGACCTCAGCGTACACGCGTCTAAGCAGCT 489	
QY	505	GTGTTAGCCCTTACTAGGAATTCGGCAGTCGAGCTCGGACAAATTTGGCAATTAGGGTTAAT 564	
Db	490	CTTCTCGGCTGTTTAAATCGGCTTGTGGCGGCTAGGGAAGTACGGGATTAGAGTCAAC 549	
QY	565	TGTTTGTCTCTTCGGGCTTCTACGCGTTTAGGCAAGAAATTTTCAGGGATTAAAT 624	
Db	550	GGCGTTGCACCATACCGGTTGGGAGCGGCGATA-----AATAGCGTGACGAGGAACG 603	
QY	625	GAAGAAGAAATTTGAGAAATGTAATAAATTTTTCGGGAAATTTGAAAGGTCCAAAATTAAT 684	
Db	604	GTGAGGATGGTGAGGAATATACGCGACGCCAGCGGATTTCTCAAAGGTGTGGCTTAG 663	
QY	685	GTTGAGGATGTTGCCAATGCACTCTTTATCTGGCTAGTGTATGAGCAAAATACGTGACT 744	
Db	664	GCTCGCCATGTGCAGAGCGGCTTGTGTTCTGCTTCGGATTCGCGTTACGTTAGC 723	
QY	745	GGACAAATCTGTTTATTCATGAGGGTTTCAGGCTTCGAA 785	
Db	724	GTCAGAAATCTGGCTTTCAGCGGGGTTATAGCGTGTAA 764	

RESULT 14
ABN98918/C
ID ABN98918 standard; DNA; 833 BP.
XX
AC ABN98918;
XX
DT 01-AUG-2002 (first entry)
XX
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 686.
XX
KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
disease; crop; thale cress; tolerance factor; insect; pathogen;
nutrition; ds.
XX
OS Arabidopsis thaliana.
XX
PN US2002023281-Al.
XX
PD 21-FEB-2002.
XX
PF 26-JAN-2001; 2001US-0770445.
XX
PR 27-JAN-2000; 2000US-178472P.
XX
(GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAMI/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHAW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
WPI; 2002-400781/43.
XX
New Arabidopsis thaliana nucleic acid for identifying homologous genes,
producing compositions that modulate the expression or function of its
encoded protein, and mapping functional regions of protein -
XX
Claim 1; SEQ ID NO 686; 49pp + Sequence Listing; English.
XX
The invention relates to an Arabidopsis thaliana nucleic acid (I)
comprising a sequence capable of hybridising under stringent conditions
to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
given in the specification or its fragment. A polypeptide (II) encoded by
(I), a transgenic plant (III) comprising an exogenous nucleic acid or a
genetically modified cell (IV) comprising an exogenous nucleic acid, is
useful for screening a candidate agent for its biological effect. (I) is
useful in identifying homologous or related genes, in producing
compositions that modulate the expression or function of its encoded
protein, mapping functional regions of the protein and in studying
associated physiological pathways. (I) is also useful for the genetic
manipulation of cells, particularly plant cells. (I) is also useful in
screening assays of various plant strains to determine the strains that
are best capable of withstanding a particular disease or environmental

CC stress. (II) and (III) are useful for screening of biologically active
 CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
 CC pathways. The screened agents are useful in improved methods of treating
 CC crops to prevent or treat disease. (II) are also useful in screening
 CC programs to identify agents that mimic or enhance the action of tolerance
 CC factors. Such agents are useful in improved methods of treating crops to
 CC enhance their tolerance to environmental stress. (I) is also useful
 CC for enhancing or inhibiting production of a biosynthetic product in a
 CC plant. (III) is useful for identifying other mediators that may induce
 CC expression of proteins of interest, for establishing the extent to which
 CC a specific insect and/or pathogen is responsible for damage to a
 CC particular plant, and for identifying other mediators that enhance or induce
 CC tolerance to environmental stress, for identifying factors involved in
 CC biosynthetic pathways of nutritional, commercial, or medicinal value and
 CC for identifying productions of nutritional, commercial or medicinal
 CC value. (IV) is useful in the study of genetic function and regulation,
 CC that affect the cellular metabolism and for screening compounds
 CC that may affect the biological function of the gene or gene products.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=999909770445.

XX at seqdata.uspto.gov/sequence.html?DocID=999909770445.
 SQ Sequence 833 BP; 193 A; 250 C; 177 G; 203 T; 10 other;
 Query Match 18.2%; Score 149; DB 24; Length 833;
 Best Local Similarity 52.2%; Pred. No. 2.1e-33;
 Matches 393; Conservative 0; Mismatches 350; Indels 12; Gaps 3;

QY 32 TAGAAGGAAAGTGGCCCTATACAGGAGGAGCCAGTGGAAATGGAGAACACACGCAA 91
 DB 833 TGGATGCGAAATCGCAATATACAGGCGAGCTAGCGGATGGAGCGGCGTTA 774
 QY 92 AACTCTTCTCCCAATGAGCAAGTGGCATGTGTGTCCTCAAGATGAATAGGTC 151
 DB 773 GGCTGTTCAGGACACAGGCTAGGTGGTCTGTGACTTTCAGAGAAAGACTTGGTC 714
 QY 152 ACTCAGTTGTCGAGGCCATTCGATTCACCTACATCCATCGATGTTACTA 211
 DB 713 AAAACGTTGCGGTTCTGTGCGGAAAGCAACAGNNNTTACCGTGGCGATGTACAA 654
 QY 212 ATGAAGAGCGGTGTAAATATCCGTCGACACAGCTTCAACCTATGGAACACTGGACA 271
 DB 653 ACGAAAGAGGTNNAAACCGCTTAAGTTCACCGTCGAAAGAGCGGAGCTTGACG 594
 QY 272 TTATGTCAGCAATGAGCAATTCGATCCCAACAGCCCGCCATCAGACGAA 331
 DB 593 TTCTCTTTNNAAACCGCGGTTATGGAAC---AGCGGGAAGCTTTCGACTTGAATC 537
 QY 332 AAGCAGACTTTGAACGCGTCTCAGTGTAAATGTAACCGAGTTTCCATGATGAGAC 391
 DB 536 TGGAAACAGTTTCAACGAAACCATGGCGGTCAACGTTCTGTGCGGTGGCTTATCAAC 477
 QY 392 ACGCAGCAGCTGTATGATTCACGACGCA---GTGGCAACATAATTCACCTGCTAGTT 448
 DB 476 ACGCTGCAGAGCCATGGTGAGAAAGCAGCGGTGGTCAATCGTATGTACGACGCG 417
 QY 449 TAAGCTCAACTATGGTGGTGGTCTTTCACATGCCCTATGTTGTTCAAGCATGCTGTGT 508
 DB 416 TGGCTCGGAGATCGGTGGTCCAGGACCTCAACGGTACAGCGGCTAAGCAGCTCTTC 357
 QY 509 TAGCCCTTACTAGGAATTCGSCAGTCAGCTCGGCAATTTGGCATAGGTTAATGTT 568
 DB 356 TCGGCTGGTAAATGCGCTTGTGGCGGGCTAGGGAAGTACGGGATAGTCAACGCG 297
 QY 569 TGCTCTCTTTCGGGCTTCCATACGCTTTAGCGAAGAAATTTTCAGGGATTAATAATGAAG 628
 DB 296 TTGCACCATAGCGGTGGCGACGCGATA-----AATAGCCGTGACGAGGAACGGTGA 243
 QY 629 AAGAAATTTGAGAAATGTAATAAATTTTCGGGAAATTTGAAGGTCCTCAAAATTAATGTT 688
 DB 242 GGATGGTGGAGAAATATAGCGACCGACCGGAGTATCTCAAGAGTGTGGTCTAAGCTC 183
 QY 689 AGGATGTTGCCAATGCAGCTCTTTATCTGCTAGTATGATGACGCAAAATACGTGAGTGCAC 748

DB 182 GCATGTGGCAGAGCGGCTTTGTTCTGCTCGGATGATTCGGTTACGTTAGCGGTC 123
 QY 749 ACAATCTGTTCAATGATGAGGGTTTCAGCGCTCTGCAA 785
 DB 122 AGAATCTGCTGTTACACGGGGTTATAGCGTCGTAA 86
 RESULT 15
 AAC50609
 ID AAC50609 standard; DNA; 774 BP.
 XX AAC50609;
 AC AAC50609;
 DT 18-OCI-2000 (first entry)
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 65475.
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; Promoter; termination sequence; ss.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX
 PN EPI033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.

Search completed: February 18, 2003, 16:41:35
Job time : 259 secs

GenCore version 5.1.4_p5_4578 Copyright (c) 1993 - 2003 CompuGen Ltd.		OM nucleic - nucleic search, using sw model	
Run on:	February 18, 2003, 15:19:44	Search time 1467 Seconds (without alignments) 9041.652 Million cell updates/sec	
Title:	US-09-673-918A-1		
Perfect score:	819		
Sequence:	1 atgcagcttcgaactgcatt.....tgtccaatatccagattctt 819		
Scoring table:	IDENTITY_NUC Gapop 10.0 , Gapext 1.0		
Searched:	16154066 seqs, 8097743376 residues		
Total number of hits satisfying chosen parameters:	32308132		
Minimum DB seq length:	0		
Maximum DB seq length:	2000000000		
Post-processing:	Minimum Match 0% Maximum Match 100% Listing first 45 summaries		
Database :	EST:.* 1: em_estba:.* 2: em_esthum:.* 3: em_estin:.* 4: em_estmu:.* 5: em_estov:.* 6: em_estpl:.* 7: em_estro:.* 8: em_hic:.* 9: gb_est1:.* 10: gb_est2:.* 11: gb_hic:.* 12: gb_est3:.* 13: gb_est4:.* 14: gb_est5:.* 15: em_estfun:.* 16: em_estom:.* 17: gb_gss:.* 18: em_gss_hum:.* 19: em_gss_inv:.* 20: em_gss_pln:.* 21: em_gss_vrt:.* 22: em_gss_fun:.* 23: em_gss_mam:.* 24: em_gss_mus:.* 25: em_gss_other:.* 26: em_gss_pro:.* 27: em_gss_rod:.*		
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	362.4	44.2	738 14 BQ870873
2	348.4	42.5	704 14 BQ875148
3	342.6	41.8	702 14 BQ873587
4	290.6	35.5	585 12 BE920368
5	288.4	35.2	676 12 BG350846
6	271.6	33.2	717 14 BM812901

belongs to contig OG_CA_Contig7026, see <http://cgpdb.ucdavis.edu/> for details.

Plate: OG10 row: E column: 14.

Location/Qualifiers

FEATURES

source

```
1..738
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone_lib="OG10E14"
/clone_lib="OG_ABCDI lettuce salinas"
/lab_host="E.coli"
/notes="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformationally made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG_LIB=OG_ABCDI lettuce salinas TAG_TISSUE=roots environmental stress TAG_SEQ=ATCTCGCGG"
BASE COUNT      221 a 147 c 175 g 195 t
ORIGIN
Query Match      44.2%; Score 362.4; DB 14; Length 738;
Best Local Similarity 70.1%; Pred. No. 3.3e-96;
Matches 502; Conservative 0; Mismatches 211; Indels 3; Gaps 1;
QY 13 ACTGCATTGCGAAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGAGCCAGTGGGA 72
DB 21 ACTAGTTCAACAAGAGGCTAGAGGAAAGTTAGCATTCACCGGAGCAGCAAGCGGT 80
QY 73 ATGGAGAAGACACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGAT 132
DB 81 ATTGGTGAGTGCATGCAAAACTATTGCTGAACATGGAGCAAAAGTAGTCAATTGCAGAC 140
QY 133 GTCAAGATGAATTAGTCTACTGATTCGAGGCGATTCGACATTCCTCAATGCCACTAC 192
DB 141 ATTCAGAACCAACTTGTGCAAGCAGTTTGTGAGGCCATAGGCACATCAATCCATATAT 200
QY 193 ATCCACTGTGATGTTACTATAGAGCGGTGTAAATATGCGTGGACACACAGTTTCA 252
DB 201 GTCCATTGTGATGTAACCAATGAAGAGATGTCAAACGCTGTAGATCTGCAGTTTCC 260
QY 253 ACCTATGGAACACTGACATTATGTTACAGCAATGACAGGAATTTCTGATCCCAACAGGCC 312
DB 261 ACATACGGAACACTTGACATCATGTTTTCATATGCGGGGATAGCAGATCTCTAAACAGCCA 320
QY 313 CGCATCATAGACACGAAAGACAGACTTTTGAACGCGTTCTCAGTGTAAATGTAAACCGGA 372
DB 321 CGTATCGTTGATATGAAGAAGACAGACTTTGAGCGGTGTGCTAGTGTCAATATACAGGT 380
QY 373 GTTTTCTCTATGATGAGCAGCAGCAGCTGTATGATTCAGCAGCAGCGTGGCAACATA 432
DB 381 GTCCTTTCTAAGCATGAACATGCTGTAGGGTTATGGTTCCAGCAGCAGCTGCGTCAATT 440
QY 433 ATTTCCACTGCTAGTTTAAGCTCAACTATGGTGGTGGTCTTCTACATGCTTATTTGGT 492
DB 441 ATATCAACGGCTAGCTTGTGCTCAAAATATTGTTGTCGAGCCTCGCATGCTTATTGGCT 500
QY 493 TCAAGCATGCTGTGTAGCCCTTACTAGCAATCTGGCAGCTCGGACATTTTGGC 552
DB 501 GCAAAACACGCTGTGCCGGTTTAACCAAAATCTTCAGTGGAACTTGGACATTTGGT 560
QY 553 ATTAGGCTTAATGTTGCTCTCTTCGGGCTCTACGGCTTTTAGGCAAGAAATTTTCA 612
DB 561 ATAGAGTCAATTTGTTGCTCTCTCTATGCAATGGTCACACCACTAGCCCAAGTTTCGTT 620
QY 613 GGGATTTAAATAGAGAAATTTGAGATGTAAATAACTTTGGCGGAAATTTGAAAGGT 672
DB 621 GGGCTTGA -- -GGGGAAGCAATTAGAAACTGTGTAACCTTACTTTGGGAATCTTAAAGGA 677
```

QY 673 CCAAAATTAAATGTTGAGATGTTGCCAATGCAGCTCTTTATCTGCTAGTGATGA 728
DB 678 CCACGCTCAGACAGATGATGTTGCTAAGGCTGCACACTTTTCTTGTGAGTGCGGA 733

RESULT 2

BO875148

LOCUS

BO875148 BO875148.1 704 bp mRNA linear EST 15-AUG-2002
DEFINITION OG17E21.yg.ab1 OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
VERSION BO875148 mRNA sequence.
KEYWORDS BO875148.1 GI:22261708
SOURCE EST
ORGANISM Lactuca sativa.

REFERENCE

AUTHORS

1 (bases 1 to 704)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://compgenomics.ucdavis.edu/>
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig OG_CA_Contig7026, see <http://cgpdb.ucdavis.edu/>
for details.
Plate: OG17. row: E column: 21.

TITLE

JOURNAL

COMMENT

source

1..704

/organism="Lactuca sativa"

/cultivar="Salinas"

/db_xref="taxon:4236"

/clone_lib="OG17E21"

/clone_lib="OG_ABCDI lettuce salinas"

/lab_host="E.coli"

/notes="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformationally made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/> TAG_LIB=OG_ABCDI lettuce salinas TAG_TISSUE=roots environmental stress TAG_SEQ=ATCTCGCGG"

BASE COUNT 215 a 139 c 165 g 185 t

ORIGIN

Query Match 42.5%; Score 348.4; DB 14; Length 704;

Best Local Similarity 70.3%; Pred No. 4.5e-92;

Matches 482; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

QY 13 ACTGCATTGCGAAGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGAGCCAGTGGGA 72

DB 21 ACTAGTTCAACAAGAGGCTAGAGGAAAGTTAGCATTCACCGGAGCAGCAAGCGGT 80

QY 73 ATGGAGAAGACACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGAT 132

DB 81 ATTGGTGAGTGCATGCAAAACTATTGCTGAACATGGAGCAAAAGTAGTCAATTGCAGAC 140

QY 133 GTCAAGATGAATTAGTCTACTGATTCGAGGCGATTCGACATTCCTCAATGCCACTAC 192

DB 141 ATTCAGAACCAACTTGTGCAAGCAGTTTGTGAGGCCATAGGCACATCAATCCATATAT 200

QY 193 ATCCACTGTGATGTTACTATAGAGCGGTGTAAATATGCGTGGACACACAGTTTCA 252

DB 201 GTCCATTGTGATGTAACCAATGAAGAGATGTCAAACGCTGTAGATCTGCAGTTTCC 260

QY 253 ACCTATGGAACACTGACATTATGTTACAGCAATGACAGGAATTTCTGATCCCAACAGGCC 312

DB 261 ACATACGGAACACTTGACATCATGTTTTCATATGCGGGGATAGCAGATCTCTAAACAGCCA 320

QY 313 CGCATCATAGACACGAAAGACAGACTTTTGAACGCGTTCTCAGTGTAAATGTAAACCGGA 372

DB 321 CGTATCGTTGATATGAAGAAGACAGACTTTGAGCGGTGTGCTAGTGTCAATATACAGGT 380

QY 373 GTTTTCTCTATGATGAGCAGCAGCAGCTGTATGATTCAGCAGCAGCGTGGCAACATA 432

DB 381 GTCCTTTCTAAGCATGAACATGCTGTAGGGTTATGGTTCCAGCAGCAGCTGCGTCAATT 440

QY 433 ATTTCCACTGCTAGTTTAAGCTCAACTATGGTGGTGGTCTTCTACATGCTTATTTGGT 492

DB 441 ATATCAACGGCTAGCTTGTGCTCAAAATATTGTTGTCGAGCCTCGCATGCTTATTGGCT 500

QY 493 TCAAGCATGCTGTGTAGCCCTTACTAGCAATCTGGCAGCTCGGACATTTTGGC 552

DB 501 GCAAAACACGCTGTGCCGGTTTAACCAAAATCTTCAGTGGAACTTGGACATTTGGT 560

QY 553 ATTAGGCTTAATGTTGCTCTCTTCGGGCTCTACGGCTTTTAGGCAAGAAATTTTCA 612

DB 561 ATAGAGTCAATTTGTTGCTCTCTCTATGCAATGGTCACACCACTAGCCCAAGTTTCGTT 620

QY 613 GGGATTTAAATAGAGAAATTTGAGATGTAAATAACTTTGGCGGAAATTTGAAAGGT 672

DB 621 GGGCTTGA -- -GGGGAAGCAATTAGAAACTGTGTAACCTTACTTTGGGAATCTTAAAGGA 677

FEATURES		Location/Qualifiers	
source		1..702	
		/organism="Lactuca sativa"	
		/cultivar="Salinas"	
		/db_xref="taxon:4236"	
		/clone="QG11P03"	
		/clone_lib="QG_ABCDI lettuce salinas"	
		/lab_host="E.coli"	
		/note="Vector: pBRcDNASFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG_LIB=QG_ABCDI lettuce salinas TAG_TISSUE=roots environmental stress TAG_SEQ=ATCTCGCGG"	
BASE COUNT		215 a	137 c 164 g 185 t 1 others
ORIGIN			
Query Match		41.8%; Score 342.6; DB 14; Length 702;	
Best Local Similarity		69.7%; Pred. No. 2.4e-90;	
Matches 478; Conservative		0; Mismatches 205; Indels 3; Gaps 1;	
Qy	13	ACTGCATTCCGAGAGGCTAGAGGAAAGTTGCCCTTATACAGGAGGAGCCAGTGGA	72
Db	19	ACTAGTTCAACAGAGAGGCTAGAGGAAAGTAGCAATTGATCACCAGGAGCAACGGT	78
Qy	73	ATTGGAGAAACACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGAT	132
Db	79	ATTGGTGAAGTGCATGCAAAACTATTGCTGAACATGGAGCAAAAGTAGTCATTGGCAGAC	138
Qy	133	GTCCAAAGATGAATTAGTGCACCTAGTTGCGAGGCCATTGGCACTTCCCAATTCACCATAC	192
Db	139	ATTCAAGACCAACTTGGTCAAGCAGTTTGTGAGGCCATAGGCACATCAAACTCCATATAT	198
Qy	193	ATCCACTGTGATGTACTAATGAAGACGCTGTTAAAAATGCCGTGCACACACAGATTCA	252
Db	199	GTCCATTGTGTATGAACCAATGAAGAAAGATGCAAAAACGCTGTAGATACTGCGATTTC	258
Qy	253	ACCTATGAAAACCTGGACATATTATTCAGCAATGCAAGAAATTTCTGATCCCAACAGGCC	312
Db	259	ACATACGGAACACTTGACATCATGTTTTCGAATGCGGGATAGCAGATCCTTRACAGCCA	318
Qy	313	CGCATCATAGACAAACGAAAGACAGACTTTGAACGCGTTCTCAGTGTAAATGTAACCGGA	372
Db	319	CGTATCGTTGATAATGAAAGACAGACTTTGAGCGGTGCTTAGTGTCAATATAACAGGT	378
Qy	373	GTTCCTCATGATGAAGCAGCAGCAGCTGTATGATTCAGCAGCGAGTGGCAACATA	432
Db	379	GTCTTCTTAAGATGAACATGCTGCTGAGGTATGTTTCCAGCAGATCTGGCTCAATT	438
Qy	433	ATTTCACCTGCTAGTTTAAGCTCAACTATGGGTGGTGGTGTCTTTCACATGCCTATTGTGT	492
Db	439	ATATCAACGGCTAGTGTGCTCAATATATTGGTGGTGCAGCCTCGCATGCTTATTTGCT	498
Qy	493	TCAAAGCATGCTGTGTAGCCCTTACTAGGAATGTCGGCAGTCGAGCTCGGACAAATTTGCG	552
Db	499	GCAAAACACGCTGTCGCCGTTTAAACAAAATCTTGCAGTGGAACTTGGACAAATTTGCT	558
Qy	553	ATTAGGTTAAATGTTGTTCTCTTCGGGCTTCCTACGGCTTTAGGCAAGAAATTTTCA	612
Db	559	ATAAGAGTCAATTTGTTGTTCTCTCTATGCAATGGTACACACCTAGCCCAAGTTTCGT	618
Qy	613	GGGATTAATAAATGAAGAAGAAATTTGAGAATGTAATAAATCTTGGCGGAAATTTGAAGGT	672
Db	619	GGGCTTGAA---GGGGAACGCAATTANAACCTGATGAACCTTACTTGGGAATCTTAGGGA	675
Qy	673	CCAAAATTTAATGTTGAGGATGTTGC	698

Qy	133	GTCCAAAGATGAATTAGTGCACCTAGTTGTCGAGGCCATTGGCACTTCCCAATTCACCATAC	192
Db	141	ATTCAAGACCAACTTGGTCAAGCAGTTTGTGAGGCCATAGCAACATCAAACTCCATATAT	200
Qy	193	ATCCACTGTGATGTACTAATGAAGCAGCTGTTAAAAATGCCGTGCACACACAGATTTC	252
Db	201	GTCCATTGTGTATGAACCAATGAAGAGATGCAAAAACGCTGTAGTACTGCAAGTTTCC	260
Qy	253	ACCTATGAAAACCTGGACATATTATTCAGCAATGCAAGAAATTTCTGATCCCAACAGGCC	312
Db	261	ACATACGGAACACTTGACATCATGTTTTCGAATGCGGGATAGCAGATCCTTRACAGCCA	320
Qy	313	CGCATCATAGACAAACGAAAGACAGACTTTGAACGCGTTCTCAGTGTAAATGTAACCGGA	372
Db	321	CGTATCGTTGATAATGAAAGACAGACTTTGAGCGGTGCTTAGTGTCAATATAACAGGT	380
Qy	373	GTTCCTCATGATGAAGCAGCAGCAGCTGTATGATTCAGCAGCGAGTGGCAACATA	432
Db	381	GTCTTCTTAAGCATGAACATGCTGCTAGGGTTATGGTTCAGCAGCATCTGGCTCAATT	440
Qy	433	ATTTCACCTGCTAGTTTAAGCTCAACTATGGGTGGTGGTGTCTTTCACATGCCTATTGTGT	492
Db	441	ATATCAACGGCTAGCTTGTGCTCAATATTTGGTGGTGCAGCCTCGCATGCTTATTTGCT	500
Qy	493	TCAAAGCATGCTGTGTAGCCCTTACTAGGAATGTCGGCAGTCGAGCTCGGACAAATTTGCG	552
Db	501	GCAAAACACGCTGTCGCCGTTTAAACAAAATCTTGCAGTGGAACTTGGACAAATTTGCT	560
Qy	553	ATTAGGTTAAATGTTGTTCTCTTCGGGCTTCCTACGGCTTTAGGCAAGAAATTTTCA	612
Db	561	ATAAGAGTCAATTTGTTGTTCTCTTATGCAATGGTGCACACCACTAGCCCAAGTTTCGT	620
Qy	613	GGGATTAATAAATGAAGAAGAAATTTGAGAATGTAATAAATCTTGGCGGAAATTTGAAGGT	672
Db	621	GGGCTTGAA---GGGGAACGCAATTAGAACTGTGATGAATCTTGGGAATCTTAGGGA	677
Qy	673	CCAAAATTTAATGTTGAGGATGTTGC	698
Db	678	CCAAACGCTCAAAACAGATGTTGC	703

RESULT 3

BQ873587

LOCUS

DEFINITION

QGI1P03.yg.abl QG_ABCDI lettuce salinas Lactuca sativa cDNA clone

QGI1P03, mRNA sequence.

Q873587

Q873587.1

GI:22260146

EST.

Q873587

Lactuca sativa.

Lactuca sativa

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.

1 (bases 1 to 702)

Q873587

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.

Lettuce and Sunflower ESTs from the Compositae Genome Project

<http://compgenomics.ucdavis.edu/>

Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]

belongs to contig QG_CA.Contig7026, see <http://cgpdb.ucdavis.edu/> for details.

Plate: QG11 row: P column: 03.

Db	676	CTAACGCTCAGAACAGATGATGTCG	701	
BE920368	585 bp	mRNA	linear	EST 02-OCT-2000
LOCUS	EST424137	potato leaves and petioles	Solanum tuberosum	cDNA clone
DEFINITION	cSTB5H14 5' sequence, mRNA sequence.			
ACCESSION	BE920368			
VERSION	BE920368.1	GI:10446444		
KEYWORDS	EST.			
SOURCE	potato.			
ORGANISM	Solanum tuberosum			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.			
AUTHORS	van der Hoeven, R.S., Bezzerides, J., Holt, I.E., Liang, F., Cho, J., Utterback, T., Hansen, C.L., Doan, B., Bougri, O., Buell, C.R., Ronning, C.M., Fry, W.E., Tanksley, S.D. and Baker, B.			
TITLE	Generation of ESTs from potato leaves and petioles			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Cathy Ronning The Institute for Genomic Research For clone request: Please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com.			
FEATURES	source			
	1..585			
	/organism="Solanum tuberosum"			
	/cultivar="Kennebec"			
	/db_xref="taxon:4113"			
	/clone="cSTB5H14"			
	/clone_lib="potato leaves and petioles"			
	/tissue_type="leaflets and petioles"			
	/dev_stage="8 weeks old plants"			
	/lab_host="SOLR"			
	/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Tissue was supplied by Dr. Fry (Cornell University). Leaflets and petioles were isolated from 8 week old greenhouse grown plants. The plants were watered and fertilized freely. The tissue was immediately frozen in liquid nitrogen."			
BASE COUNT	164 a 154 c 139 g 128 t			
ORIGIN				
	Query Match 35.5%; Score 290.6; DB 12; Length 585;			
	Best Local Similarity 71.6%; Pred. No. 5.8e-75;			
	Matches 396; Conservative 0; Mismatches 154; Indels 3; Gaps 1;			
QY	17	CATTTCGCAAGAAGGCTAGAGAAAGTTGGCCCTTATAACAGGAGGAGGACGATGGAATTG	76	
Db	33	CAATCGGCAAAAGGCTAGAGGAAAGTAGCAATTGTAAACGGGAGGCTAGTGGTATCG	92	
QY	77	GAGAACCCACAGCAAACTCTCCCAACATGAGGCAAAAGTGGCCATGCTGATGTCC	136	
Db	93	GTGAACCAATTGCAAGCTCTCTGAACATGGAGGCAAAAGTGTGTCGGATGTCC	152	
QY	137	AGATCAATTAAGTTCACATCTTCGAGGCCAT---TGGCACTTCCATTCACCTACA	193	
Db	153	AGATGAATCGGCAACTCATGTCATCAACGCCCTCGGAGGCTCATCCAATCCATTTACA	212	
QY	194	TCCACTGTGATGTTACTAATGAACAGCGTGTAAATGCCGTGGACACACAGTTTCAA	253	
Db	213	TCCACTGTGATGTTACTAATGAACAGCGTGTAAATGCCGTGGACACACAGTTTCAA	272	
QY	254	CCTATGGAAAACCTGACATATTGTTGACAAATGACAGGAATTTCTGATCCACAGGCCCC	313	
Db	273	CATTGGAAAACCTGACATATTGTTGACAAATGACAGGAATTTCTGATCCACAGGCCCC	332	
QY	314	GCATATAGACAAACGAAAACGACAGCTTTGAACGCGTCTCTAGTGTAAATGTAACCGGAG	373	
Db	333	GAATCATAGACAAACGAAAACGACAGCTTTGAACGAGTCCCTTAGCATCAACAGTCACAGGAG	392	

Db 268 CCCTCCGAGCTCACAAAGACACACGTCCTCAAGAAAGCTGTGCGACAGAACCACTCCCCAC 327

QY 255 CTATGAAAACTGGACATTATGTCAGCAATGACAGGAATTTCTGATCCCAACAGGCCCG 314

Db 328 ATTGGAACCTGACATCATGCTGCAATGCTGTCATATGCGAGGACCAAGTCAAG 387

QY 315 CATCATAGACAGGAAAGACAGACTTTGAACGCGTCTCTCAGTGTAATGTAACCGGAGT 374

Db 388 AATCATAGACAAACAGAAAGAGATTTCAAGAGAGTCCCTTTAGCATCAAGCTAACAGGAGT 447

QY 375 TTCTCTATGATGACAGCAGCAGCTGTATGATTTCCAGCAGCGAGTGGCAACATAAT 434

Db 448 TTCTTGACATGAGCATGCGCTCGCTCATGTCGCGAGCGGTAGCGGTGCATCAT 507

QY 435 TTCCACTGCTAGTTTAAGCTCAACTATGGTGGTGGTCTTCCACATGCGCTATTTGGTTC 494

Db 508 CTCACCGCTAGTGTAAAGCTCCAGCGTCGAGCTGCAGCATCCCATGCTACTGCAAGCTC 567

QY 495 AAAGCATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAAATTTGGCAT 554

Db 568 CAACACGCTGTGTGGGGGTCAACGAACCTAGCGGTGGAGCTAGGACAATTCGGTAT 627

QY 555 TAGGGTAAATTTGTTCTCTCTCTTGGGCT 584

Db 628 ACGTGCAATGCTGTCAACATACCGAT 657

RESULT 6

LOCUS BM812901

DEFINITION EST590994 BNIR Medicago truncatula/Meloidogyne incognita mixed EST library cDNA clone pBNIR-4010, mRNA sequence.

ACCESSION BM812901

VERSION 1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Medicago truncatula/Meloidogyne incognita mixed EST library.
Medicago truncatula/Meloidogyne incognita mixed EST library.
Eukaryota; mixed EST libraries.
1 (bases 1 to 717)
Bird, D., Koltai, H., Samac, D., Town, C.D., Van Aken, S., Utterback, T.,
Cheung, F., Tsai, J., and Fraser, C.M.
ESTs from roots of Medicago truncatula after infection with the
nematode Meloidogyne incognita
Unpublished (2002)
Plant Nematode Genetics Group
North Carolina State University
Box 7616, Raleigh, NC 27695, USA
Tel: 919 515 6813
Fax: 919 515 9500
Email: david.bird@ncsu.edu
TIGR sequence name: MTQAB89TK More information is available at:
www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gat CC).

FEATURES

source

1. .717
/organism="Medicago truncatula/Meloidogyne incognita mixed
EST library"
/cultivar="A17"
/db_xref="taxon:188702"
/clone="pBNIR-4010"
/clone_lib="BNIR"
/tissue_type="3 week old roots"
/dev_stage="3 days after infection with Meloidogyne
incognita second stage larvae"
/lab_host="XL0LR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0LR cells."

BASE COUNT 202 a 131 c 179 g 205 t

ORIGIN

Query Match 33.2%; Score 271.6; DB 14; Length 717;

Best Local Similarity 62.1%; Pred. No. 2.7e-69;

Matches 452; Conservative 0; Mismatches 264; Indels 12; Gaps 1;

QY 37 GGAAGAATGCCCCCTTATAACAGAGAGGAGCCAGTGGAAATTTGGAGAAACACAGCAAAACTC 96

Db 2 GGTACGATAGCTTTGATCACCGTGTGGAGCGGATTTGGTGGAGCCACTGCAAGACTC 61

QY 97 TTCTCCCAACATGGAGCCCAAGTTGCCATTGCTGATGTCTCAAGATGAATTAGTCTACTCA 156

Db 62 TTCTCCGAGCATGGAGCTCAAGTCGTTAGCTGACATCCAAAGCAGCAAAAGGACACTCT 121

QY 157 GTTGTGCGAGCCATTGGCACCTTCCAACTTACCTACATCCACTGTCATGCTTACTAATGAA 216

Db 122 ATTTGCAAGGAATTACAAAATCGTCTTCTTCTATGTTGTCGACGTGACAAAGGAA 181

QY 217 GACGGTGTAAATAATGCGGTGACACACAGTTTCAACCTATGGAATACTGGACATATTATG 276

Db 182 GAAGACATTTGAAAATGCCGTGAACACACACCGTTTCAAGTACGGTAAACTAGATATCATG 241

QY 277 TTCAGCAATGCAAGGAATTTCTGATCCCAACAGCCCGCATCATACACAAAGAAAGCA 336

Db 242 TTTAACACGCTGGCATAGCGGTGTAAACAAACAAAGATACTTTGAAAACAAATTTATCT 301

QY 337 GACTTTTGAACGGTCTCTCAGTGTAAATGTAACGGGAGTTTTCCTATGTCATGAAGCACGCA 396

Db 302 GAATTTGAGGATGTGATTAAGTAAACTTGACTGGCGCTCTTCTTGGAAACAAACATGCA 361

QY 397 GCACGTGTTTATGATTCACAGCAGCAGTGGCAACATAAATTTCCACTGCTAGTTTAAAGCTCA 456

Db 362 GCAAGGGTAATGATCCCTCTCGACGGGTAGCATAAATTAACACACAGCTAGTGTGCTGGA 421

QY 457 ACTATGGGTGGTGGTCTTCACATGCCCTATTGTTGTTTCAAGCATGCTGTGTAGCCCTT 516

Db 422 AGCATTTGGTGGTGTGCTCCCATGTACAAAGTTCAAGCAGCCCGCTTGTGGGCTG 481

QY 517 ACTAGGAATCTGGCAGTCTGAGCTCGGACAAATTTGGCATTTAGGTTAATTTGTTGTCCTCT 576

Db 482 ATGAGAAACACGGCTATCGAGCTTGGGCCATTCGGGTGAATTTGTGTGTCGCT 541

QY 577 TTCGGGCTTCTACAGGCTTTAGGCAAGAAATTTTCAGGATTAATAAATGAAGAAGAAATTT 636

Db 542 TATTTCTTTCGCGACTCCGATGGTTACAAATTTCTTTAAGCTTATGATGTTGTTGAGCTCA 601

QY 637 GAGATGTAATAAATTTTCGGGGAATTTGAAAGGTTCCAAATTTAATGTTGAGGATGTT 696

Db 602 GATATTTTCTTAACCTAAAGGTTACTAATCTTGTGCCAAAA-----GATGTG 649

QY 697 GCAATGCGAGCTCTTTATCTGCTAGTGTAGGCAAAATACGTGAGTGACACAAATCTG 756

Db 650 GCTGAAGCTGCTTTGTTATTTGGGAAGTGATGAGCTCTAAGTATGTTAGTGGTCTTAATCTT 709

QY 757 TTCATTGA 764

Db 710 GTCATAGA 717

RESULT 7

LOCUS B0630495

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

637 bp mRNA linear EST 02-JUL-2002

saq08b06.y1 Gm-cl045 Glycine max cDNA clone SOYBEAN CLONE ID: 5'

similar to TR:P93697 P93697 CPD12 PROTEIN. ;, mRNA sequence.

B0630495

BQ630495

GI:21678144

EST.

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Best Local Similarity 62.1%; Pred. No. 2.2e-65; Matches 431; Conservative 0; Mismatches 251; Indels 12; Gaps 1;

QY 68 GTGGAATTGGAGAACACAGCAAACTCTCTCCCAACATGAGCAAAAGTTGCCATTG 127
Db 4 CGGCAATTGGTGGAGCACTGCAAGACTCTCTCCGAGCATGAGCTCAAGTGTATAG 63

QY 128 CTGATGTCCAAGATGAATAGTCTACTAGTTGTGCGAGCCATTGGCACTTCCAATTCCA 187
Db 64 CTGACATCCAAGACGCAAAAGGACACTCTATTGCAAGGAATTTACAAAATCGTCTCTT 123

QY 188 CCTACATCCACTGTGATGTTACTAATGAGACGGTGTAAAAATGCCGTGGACAACACAG 247
Db 124 CCTATGTTGTTGTCAGTGCACAAAGGAAGAGACATTTGAAAATGCCGTGAACACACAC 183

QY 248 TTTCAACCTATGAAAACTGGACATTATGTTTCAGCACTCAGCAATTTCTGATCCCAACA 307
Db 184 TTTTCAAGTACGTAACATAGATATCATGTTTAAACAACGCTGCATTAAGTAAACTTGA 243

QY 308 GGGCCCGCATCATAGCAACGAAAGAGACACTTTGAAAGCGCTTCCTCAGTGAATGTAA 367
Db 244 AAACAAGATACTTGAACAAATATCTGAATTCAGGATGTGATTAAGTAAACTTGA 303

QY 368 CCGGAGTTTTCCTATGCATGCAACGACGACGACCTGTTATGATTCAGCAACGAGTGCCA 427
Db 304 CTGGCTCTTCTTGGAAACAAACATGCAAGGAGGTAATGATCCCTGTCGACGGGGTA 363

QY 428 ACATAATTTCCACTGCTAGTTTAAAGCTCAACTATGTTGGTGGTCTTCCATGCTCTATT 487
Db 364 GCATATTAACACAGCTAGTGTGGGAGCAATGTTGGTGTGCTCCATGCTCTACA 423

QY 488 GTGGTCAAGCATGCTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAA 547
Db 424 CAAGTTCAAGACGCGCTGTTGGGCTGATGAGAAACACGCGTATCGAGCTTGGGCCAT 483

QY 548 TTGGCATTAGGTTAATGTTGTCCTCTTTCGGGCTTCTACGGCTTCTAGGCAAGAAAT 607
Db 484 TCGGTTATTCGGGTTGAATGTGTGTCGCCCTTATTTCTGCGACTCGGAGTGGTTACAAAT 543

QY 608 TTTCAAGGATTAATAAGAGAAATTTGAGATCTTAATAACTTTGCGGAAATTTGA 667
Db 544 TCTTTAAGCTTGATGTTGGAGCTCTAGATATTTTCTACCTAAAGGTTACTAATC 603

QY 668 AAGTCCAAATTAATGTTGAGGATGTTGCCAATCGACTCTTTATCTGGTGTAGTATG 727
Db 604 TTGTGCCAAA-----GATGTGCTGAAGCTGCTTTGTATTTGGGAAGTATG 651

QY 728 AGCAAAATACGTGAGTGACACAACTCTGTTCA 761
Db 652 AGCTTAAGTATGTTAGTGTCTTAATCTTGTCA 685

RESULT 9
LOCUS BG456805
DEFINITION NF098D03PL1F1028 Phosphate starved leaf Medicago truncatula cDNA clone NF098D03PL 5', mRNA sequence.
ACCESSION BG456805
VERSION BG456805.1 GI:13380130
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 682)
AUTHORS Liu, J., Scott, A. D., Harris, A. R., Gonzales, R. A., Bell, C. J., Flores, H. R., Inman, J. T., Weller, J. W., May, G. D. and Harrison, M. J.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula phosphate-starved leaf library
JOURNAL Unpublished (2000)
COMMENT Contact: Harrison MJ

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel.: 580 221 7325
Fax: 580 221 7380
Email: mjharrison@noble.org
Insert Length: 682 Std Error: 0.00
Plate: 098 row: D column: 03
Seq primer: TCACACAGAAACAGCTATGAC.
Location/Qualifiers
1. 682
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF098D03PL"
/clone_lib="Phosphate starved leaf"
/tissue_type="leaf"
/dev_stage="trifoliolate"
/note="Vector: Lambda Zap; At the trifoliolate stage, M. truncatula plants were transplanted to phosphate-free sand and grown for a further 30 days. During this 30 day period, the plants were fertilized twice weekly with 1/2 Hoaglands solution containing only 20mM potassium phosphate. RNA was prepared from above ground tissues."
BASE COUNT 210 a 151 c 150 g 169 t 2 others
ORIGIN
Query Match 31.5%; Score 257.6; DB 12; Length 682;
Best Local Similarity 63.4%; Pred. No. 3.8e-65;
Matches 392; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 14 CTGCATTCGCAAGAGGGCTAGAGGAAAGTTGCCCTTATAACAGGAGGAGCCAGTGGAA 73
Db 59 CTGCTGCCCTAAGAGGCTTGAAGGAAAGTGGCTCTGATCACCCTGTTGCAAGCGGCA 118

QY 74 TTGGGAAACACGACAAACTCTCTCCACATCGAGCCAAAGTTGCCATTCGTGATG 133
Db 119 TCGGGAGGCCCGGCAAGACTCTTCTCCACACGCGAGCTCAAGTCGTTATAGCTGATA 178

QY 134 TCCAAGATGAATTAGTCTACTGATGTTGCGAGGCCATTCGCAATTCACCTACA 193
Db 179 TCCAAGATGACATAGTGTACTCTATTGCGCAGAGTTACACAANTCTCTGTCACCTATG 238

QY 194 TCCACTGTGATGTTACTAATGAAGCGGTGTTTAAAAATCCGTGGACACACAGTTTCAA 253
Db 239 TCATTGCGACGTAAACAAAGGAAAGACATTGAAAACGCCGTAAACACACACCGTTTCCA 298

QY 254 CCTATGGAAAACCTGGACATTATGTTCCAGCAATGCAGGAATTTCTGATCCCAACAGGCC 313
Db 299 AACACGGTAAATTAGATATCATGTTTAAACACGCGCATAAACCCGGTATAACAAACAA 358

QY 314 GCATCATAGACAAACGAAAGCAGACTTTGAACGCCGTTCTCAGTGTAAATGTAACCGGAG 373
Db 359 ACATACTTGAACAACAATTTATCCGAATTTCAAGAAGTGATTGACATTAACCTGACCGGTG 418

QY 374 TTTTCTGATGCATGAAGCAGCAGCAGCTGTTTATGATTCAGCAGCAGCTGCGCAACATAA 433
Db 419 TATTTCTCGGAACAAACACCGCTGCTAGGTAATGACCCCGTTTCAGCAGGAGTAGCATAA 478

QY 434 TTTCCACTGCTAGTTTAACTCACTATGATGGTGGTGGTCTTTCACATGCTATTGTTGTT 493
Db 479 TTAACACAGCTAGTGTGTTTGGTATGATAGTGTGCTTACATGCTTACATGCTACACAAGTT 538

QY 494 CAAAGCATGCTGTTGTTAGCCCTTACTAGGAATCTGCGAGCTCGAGCTCGGCAATTTGGCA 553
Db 539 CAAAGCAGCCCGTGTGTTGGCTAATGAAAATACNGCCATCGAGCTTGGACCGCTACGGTA 598

QY 554 TTAGGTTTAATTTTGTCTCTCTCTTTCGGGCTTCTCTACGGCTTTTGTAGCAAGAAATTTTCAG 613
Db 599 TTCGAGTGAATTTGTGTGTCACCTTATGTTGTCGGAAGTCCGTTGGCTGAANAATTTCTTTA 658

QY 614 GGATTTAAATATGAGAAG 631
Db 659 AGCTTGATGATGGAG 676

JOURNAL COMMENT	Unpublished (2000) Contact: Deborah A. Samac Department of Plant Pathology University of Minnesota 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA Tel: 612 625 1243 Fax: 651 649 5058 Email: debbys@puccini.crl.umn.edu University of Minnesota name: M278844e TIGR sequence name: MTFDH237K More information is available at: http://chrysie.tamu.edu/medicago Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
FEATURES	Location/Qualifiers 1..681 /organism="Medicago truncatula" /cultivar="genotype AL7" /db_xref="taxon:3880" /clone="pDSIL-43C22" /tissue_type="leaves infected with Colletotrichum trifolii" /dev_stage="cotyledons and primary leaves harvested 5 and 8 days after inoculation with Colletotrichum trifolii" /lab_host="E. coli strain XL0R" /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from cotyledons and primary leaves harvested 5 and 8 days after inoculation with Colletotrichum trifolii. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XL0LR cells. Note: EST may be of fungal origin."
source	195 a 126 c 168 g 192 t
BASE COUNT	
ORIGIN	
Query Match	31.4%; Score 257.4; DB 12; Length 681;
Best Local Similarity	62.0%; Pred. No. 4.4e-65;
Matches	430; Conservative 0; Mismatches 251; Indels 12; Gaps 1;
QY	52 ATACAGGAGGCCAGTGGAAATGAGAAACACAGCAAACTCTTCCCAACATGGA 111
Db	1 ATACCCGGTGGTCCGAGCGCATGTGTGGCCACTGCAGACACTCTTCCGAGCATGA 60
QY	112 GCCAAAGTTGCCATGCTGATGCCAAGATGAATTAGGTCACCTCAGTTGTGCGAGGCATT 171
Db	61 GCTCAAGTCGTTATAGCTGACATCCAGACGACAAGGACACTCTATTTCGAAGGAATTA 120
QY	172 GGCATTCCCAATTCACCTACATCCACTGTGATGTTTACTAATGAAGACGGTGTAAAAAT 231
Db	121 CAAAAATCGTCTTCTTCCATGTTCGTTGACGTGACAAAGGAAGACATTTGAAAT 180
QY	232 GCGGTGGACACACAGTTTCAACCTATGAAAACTGGACATATATGTCAGCAATGCAGGA 291
Db	181 GCGGTGAACACACACCGTTTTCAGTAGTCGGTAAACTAGATATCATCTTTTAACACGCTGC 240
QY	292 ATTCTGTATCCCAACAGGCCGCCATCATAGACAAGCAAAAGACACTTTTGAACCGGTT 351
Db	241 ATAAGCGGTGTAACAACAACAAGACTATTGAAACAATATTCTGAATTTGAGGATGTG 300
QY	352 CTCAGTGTAAATGTAACCGGAGTTTTCCTATGCATGAAGCAGCAGCAGCTGTCTATGATT 411
Db	301 ATTAAGTAAACTTGACTGGCGTCTTCTTGGAAACAACAATGCAGCAAGGGTAATGATC 360
QY	412 CCAGCAGCAGTGGCAACATAATTTCCACTGCTAGTTTAAGCTCACTATGGTGGTGGT 471
Db	361 CTGCTCGACGGGTAGCAATAATTAACACAGCTAGTGTGGTGGAAAGCATTTGGTGGTTGT 420
QY	472 TCTTCATCGCTTATTGTGGTCAAGAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCA 531
Db	421 GCTCCACATGCCTACCAAGTTCAAGACACGCGGTTGTTGGGCTGATGAGAAACACGCT 480
QY	532 GTCGAGCTCGGACAATTTGGCATTAGGTTAATGTTTGTCTCTTTCGGGCTTCCTACG 591
Db	481 ATCGAGCTTGGGCCAATTCGGTATTCGGTGCAATGTGTGCGCTTATTCTTCTGGCACT 540
QY	592 GCTTTAGGCAAGAAATTTTCAGGGATTAAAAATGAAGAAGATTTTGAGATGTAAATAAC 651
Db	541 CGGATGTTACAAATTTCTTTAAGCTTGATGATGGTGGAGCTCTAGATATATTTTCTAAC 600
QY	652 TTTGCGGGGAAATTTGAAAGGTCACAAATTAATGTTTGAGGATGTTGCCAATGCAGCTCT 711
Db	601 CTAAGGGTACTAATCTTGTGCCAAAA-----GATGTGCTGAAGCTGCTTTG 648
QY	712 TATCTGGCTAGTGATGAGGCAAAATACGTGAGT 744
Db	649 TATTTGGGAAGTGATGAGTCTAAGTATGTTAGT 681
RESULT 12	
BO862029	
LOCUS	BO862029 695 bp mRNA linear EST 14-AUG-2002
DEFINITION	OGC1h09_yq.ab1 OG-ABCDI lettuce salinas Lactuca sativa cDNA clone
ACCESSION	OGC1h09 mRNA sequence.
VERSION	BO862029
KEYWORDS	BO862029.1 GI:22247494
SOURCE	EST.
ORGANISM	Lactuca sativa. Lactuca sativa. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.
REFERENCE	1 (bases 1 to 695)
AUTHORS	Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K. Lettuce and Sunflower ESTs from the Compositae Genome Project http://compgenomics.ucdavis.edu/ Unpublished (2002)
TITLE	Contact: Alexander Kozik [R.W.Michelmore] Department of Vegetable Crops, R.W.Michelmore Lab University of California at Davis (UCD) Asmundson Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659 Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu] singleton, see http://cgdb.ucdavis.edu/ for details. Plate: QGCI row: h column: 09.
JOURNAL COMMENT	Location/Qualifiers 1..695 /organism="Lactuca sativa" /cultivar="Salinas" /db_xref="taxon:4236" /clone="OGC1h09" /lab_host="E.coli" /note="Vector: pBRCNDSFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/ TAG_LIB=OG-ABCDI lettuce salinas TAG_TISSUE=chemical induction TAG_SEQ=GTGAGCCGG"
BASE COUNT	208 a 136 c 163 g 186 t
ORIGIN	
Query Match	31.3%; Score 256.6; DB 14; Length 695;
Best Local Similarity	61.4%; Pred. No. 7.6e-65;


```

ACCESSION      BQ982959
VERSION        BQ982959.1  GI:22400484
KEYWORDS
SOURCE        Lactuca sativa.
ORGANISM      Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE      1 (bases 1 to 544)
AUTHORS        Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
                Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
                ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
                Church,S., Jackson,L. and Bradford,K.
TITLE          Lettuce and Sunflower ESTs from the Compositae Genome Project
                http://compenomics.ucdavis.edu/
JOURNAL        Unpublished (2002)
COMMENT        Contact: Alexander Kozik [R.W.Michelmore]
                Department of Vegetable Crops, R.W.Michelmore Lab
                University of California at Davis (UCD)
                Asmundson Hall, UCD, Davis, CA 95616, USA
                Tel: 1-(530)-742-1742
                Fax: 1-(530)-752-9659
                Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
                belongs to contig QG_CA.Contig3351, see http://cgpdb.ucdavis.edu/
                for details.
                Plate: QGE17 row: N column: 08.
FEATURES       Location/Qualifiers
                1..544
                /organism="Lactuca sativa"
                /cultivar="L.serriola"
                /db_xref="taxon:4236"
                /clone.lib="QG.EFGHJ lettuce serriola"
                /lab.host="E.coli"
                /note="Vector: pBRCDNASFIAB; The library was constructed
                from 10 different sources of RNA from a single genotype.
                Separate cDNAs were generated using primers that
                incorporated unique 5' and 3' tags to distinguish each
                source of RNA. cDNAs were then pooled, size-fractionated,
                directionally cloned into a custom medium-copy vector and
                transformations made with four size classes to minimize
                size bias. Details of each source of RNA and library
                construction can be obtained at http://cgpdb.ucdavis.edu/
                TAG.LIB-QG.EFGHJ lettuce serriola
                TAG_TISSUE=leaves dark grow
                TAG_SEQ=GTAGTCGGG"
BASE COUNT     159 a 95 c 123 g 167 t
ORIGIN
Query Match      30.8%; Score 252.6; DB 14; Length 544;
Best Local Similarity 67.6%; Pred. No. 1e-63;
Matches 370; Conservative 0; Mismatches 174; Indels 3; Gaps 1;
QY 259 GGAAACTGGACATTATGTTACGAAATTCGAGTAATTCGATCCCAACAGCCCGCAATC 318
Db 1 GGAATACTTGATATCATGTTTGTATGTCAGGAATAATCGATCCTTAACAAGGATCACATC 60
QY 319 ATAGCAACGAAAGACACACTTGAACCGGTTCTCAGTGTAAATGTACCCGGAGTTTC 378
Db 61 ATCGATAATGAAATAATTCGATTGTAACCTGGCTAGTGTCAACGTCACAGGTGCTCTTT 120
QY 379 CTATGATCAAGCAGCAGCGTGTATGATTCAGACGACGAGTGCGCAACATAATTTCC 438
Db 121 CTAAGCATGAACATGACGCTAGGTTATGTTTCCACACAGAGCTGGGTCAATATATCA 180
QY 439 ACTGCTAGTTTAAAGCTCAACTATGGTGGTGGTCTTCTACATGCCTATTTGGTTCAAG 498
Db 181 ACAGCTAGTGTGTGCAAAATATCGGTGGCTTGTGCCACACATGCTTACACTTTGTCAAAG 240
QY 499 CATGCTGTGTTAGCCCTTACTAGGATATCGGAGTCGAGCTCGGACAAATTTGGCATTAGG 558
Db 241 CATGCTGTAGCTGGTGTAAACCAAGAATCTTGCAGTTGAGCTTGGACAATTTGGTATCGA 300

```

```

QY 559 GTTAATTTGTTCTCCTTTTCGGGCTTCCTACGGCTTTTAGCAAGAAATTTTCAGGGATT 618
Db 301 GTCAATTGCTTTTACCTTTACGCAATAGATACACCAAGCCACAGGTTTCACGGGTCTT 360
QY 619 AAAATGAAGAAGAAATTTGAGAATGAATAAACTTTTGGGGGAAATTTGAAGAGTCCAAA 678
Db 361 AAA---GGGGAGGATTTTCAAGAACAGAGGTAACACTCAGTTTGCAACCTTAAGGGTGTGACA 417
QY 679 TTTAATGTTGAGGATGTTCCCAATGCAGCTCTTTATCTCGCTAGTCAGCAAAATAC 738
Db 418 CTTACAACAGATGATGTTCTTAAGCTGCTGTTTTTTTGGCGAGTGATAGGCTAAGTAT 477
QY 739 GTGAGTGGACAAATCTGTTTCAATGATGAGGGTTCAGCGTCTGCAATTTCTGTAATCAA 798
Db 478 ATTAGTGGACATAATTTGTTTATTGATGATGATTTAGCATTTGGTAACCCATCATTTCAAT 537
QY 799 GTGTTCC 805
Db 538 ATGTTTC 544

RESULT 15
BQ914508
LOCUS
DEFINITION
BQ914508
ACCESSION
VERSION
BQ914508.1  GI:22313289
KEYWORDS
SOURCE
ORGANISM
common sunflower.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
Heliantheae; Helianthus.
REFERENCE      1 (bases 1 to 715)
AUTHORS        Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
                Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
                ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
                Church,S., Jackson,L. and Bradford,K.
TITLE          Lettuce and Sunflower ESTs from the Compositae Genome Project
                http://compenomics.ucdavis.edu/
JOURNAL        Unpublished (2002)
COMMENT        Contact: Alexander Kozik [R.W.Michelmore]
                Department of Vegetable Crops, R.W.Michelmore Lab
                University of California at Davis (UCD)
                Asmundson Hall, UCD, Davis, CA 95616, USA
                Tel: 1-(530)-742-1742
                Fax: 1-(530)-752-9659
                Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
                singleton, see http://cgpdb.ucdavis.edu/ for details.
                Plate: QHB10 row: W column: 20.
FEATURES       Location/Qualifiers
                1..715
                /organism="Helianthus annuus"
                /cultivar="RHA801"
                /db_xref="taxon:4232"
                /clone.lib="QH_ABCDI sunflower RHA801"
                /lab.host="E.coli"
                /note="Vector: pBRCDNASFIAB; The library was constructed
                from 11 different sources of RNA from a single genotype.
                Separate cDNAs were generated using primers that
                incorporated unique 5' and 3' tags to distinguish each
                source of RNA. cDNAs were then pooled, size-fractionated,
                directionally cloned into a custom medium-copy vector and
                transformations made with four size classes to minimize
                size bias. Details of each source of RNA and library
                construction can be obtained at http://cgpdb.ucdavis.edu/
                TAG.LIB-QH_ABCDI sunflower RHA801
                TAG_TISSUE=callus
                TAG_SEQ=GTACTCGGG"

```

BASE COUNT		218 a	140 c	161 g	196 t		
ORIGIN							
Query Match		30.8%; Score 252.6; DB 14; Length 715;					
Best Local Similarity		65.9%; Pred. No. 1.2e-63;					
Matches 366; Conservative		0; Mismatches 189; Indels 0; Gaps 0;					
QY	23	CAAGAAGCTAGAGGAAGTTGGCCCTTATAACAGAGGAGGCCAGTGGAAATTTGGAGAAA	82				
Db	94	CACCAAGATTGGCAGTAGGTAGCAATTAACAGAGCGCGCTCAAGGAATAGAGCGCGT	153				
QY	83	CCACAGCAAAACTCTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGTGATGCCAAGATG	142				
Db	154	CAACTGCAAAACTATTCCGTAAACATGGGCGGAAGTTATTATTGCCGATATCCAAGATG	213				
QY	143	AATTAGTCACTCAGTTGTCGAGGCCATTGGCACTTCCAAATTCACCTACATCCACTGTG	202				
Db	214	ATTTGGGTCAATCTGTTGTGAAGAAATAGGCCCTTGAAATGCATCTTATGTCCCACTGTG	273				
QY	203	ATGTTACTAATGAAGACGGTGTAAATAATGCCGTGGACACACAGTTTCAACCTTATGGAA	262				
Db	274	ATGTAACCTGTTGAGGCTGATGTAGAAAATGTTGTAACCTTTGCCACACAAGATATGGTA	333				
QY	263	AACTGGACATTATGTTACGAATTCAGGAATTTCTGATCCCAACAGGCCCGGCATCATAG	322				
Db	334	AACATAGATATTATGATCAACAATGCCGCTATTCTAGGTCAAGCGAACTCAGCATTTCTG	393				
QY	323	ACACGAAAGACAGACTTTGACCGGTCTCAGTGTAAATGTAACCGGAGTTTTCCTAT	382				
Db	394	ACACGCAAAAGCTGATTTTGATCGTTGTAAACGTGAATCTAGCGCGGTGTTCCCTG	453				
QY	383	GCATGAAGCAGCAGCAGCTGTTATGATTCACAGCAGCGAGTGGCAACATAATTTCCACTG	442				
Db	454	GAATGAAGCAGCAGCTGCTGCGCCATGATTCACAGCGCTTGTGGTAGCATATCTCGATTG	513				
QY	443	CTAGTTTAAAGTCAACTATGGGTGGTGTCTTCTACATGCCTATTGTGGTTCAAAGCATG	502				
Db	514	GGAGTGTCTTCTGTGTATCGGCGGTGTTTCCCGCACACGCGTATACGAGTACCAAACTG	573				
QY	503	CTGTGTTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGGCATTTAGGGTTA	562				
Db	574	CAATTTTGGACTCACCAGAATGTGGCGCAGAGCTTGGCAAAACATGGAATTCGTGTCA	633				
QY	563	ATTGTTTGTCTCCTTT	577				
Db	634	AATGTGTATCACCAT	648				

Search completed: February 18, 2003, 17:45:38
Job time : 1475 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 16:17:44 ; Search time 81 Seconds
(without alignments)
3100.843 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgcatt.....tgtccaataatccagattct 819

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.8	14.0	1187	1	US-08-440-856A-2
2	100	12.2	1288	1	US-08-440-856A-9
3	97.2	11.9	1236	1	US-08-440-856A-1
4	70.2	8.6	771	4	US-09-504-358-13
5	70.2	8.6	771	4	US-09-504-314-13
6	70.2	8.6	10629	4	US-09-504-358-15
7	70.2	8.6	10629	4	US-09-504-314-15
8	60.6	7.4	732	4	US-09-239-052-1
9	59.6	7.3	19702	4	US-08-961-527-7
10	57.8	7.1	1196	4	US-08-858-207A-8
11	54.4	6.6	873	4	US-09-134-001C-1502
12	54	6.6	774	4	US-09-134-001C-725
13	52.8	6.4	2774	4	US-09-363-189B-5
14	51.2	6.3	825	4	US-09-134-001C-1594
15	50.4	6.2	786	4	US-09-468-738A-28
16	50.4	6.2	786	4	US-09-940-019-28
17	49.2	6.0	750	4	US-09-134-001C-1988
18	45.4	5.5	4403765	4	US-09-103-840A-2
19	43.6	5.3	804	4	US-09-134-001C-2205
20	42.6	5.2	1995	4	US-08-936-165A-250
21	41.6	5.1	852	4	US-09-134-001C-2509
22	40	4.9	756	3	US-08-822-322-7
23	40	4.9	756	4	US-09-466-109-7
24	39.6	4.8	650	4	US-09-468-738A-13
25	39.6	4.8	650	4	US-09-940-019-13
26	39.6	4.8	879	4	US-09-468-738A-1
27	39.6	4.8	879	4	US-09-940-019-1

28	39.6	4.8	891	4	US-09-468-738A-22	Sequence 22, Appl
29	39.6	4.8	891	4	US-09-940-019-22	Sequence 22, Appl
30	39.6	4.8	1787	4	US-09-468-738A-17	Sequence 17, Appl
31	39.6	4.8	1787	4	US-09-940-019-17	Sequence 17, Appl
32	39	4.8	852	4	US-09-367-012-2	Sequence 2, Appl
33	39	4.8	852	4	US-09-777-157A-2	Sequence 2, Appl
34	39	4.8	1303	4	US-09-367-012-8	Sequence 8, Appl
35	39	4.8	1303	4	US-09-777-157A-8	Sequence 8, Appl
36	37.8	4.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl
37	36	4.4	1155	3	US-08-793-035-2	Sequence 2, Appl
38	36	4.4	1185	3	US-08-793-035-1	Sequence 1, Appl
39	35.8	4.4	947	4	US-09-347-878-49	Sequence 49, Appl
40	35.8	4.4	981	4	US-08-815-225-1	Sequence 1, Appl
41	35.4	4.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl
42	35.2	4.3	31728	4	US-09-453-702B-64	Sequence 64, Appl
43	34.8	4.2	741	3	US-09-238-481-1	Sequence 1, Appl
44	34.8	4.2	741	4	US-09-572-810A-1	Sequence 1, Appl
45	34.4	4.2	777	4	US-09-504-358-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-440-856A-2
; Sequence 2, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELLAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE. N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,856A
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLMAN, ROBERT A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 05463-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1517
; TELEFAX: (202) 887-0763
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-440-856A-2

Query Match	14.0%	Score 114.8;	DB 1;	Length 1187;
Best Local Similarity	52.0%;	Pred. No. 7.3e-27;		
Matches 313;	Conservative	0;	Mismatches 277;	Indels 12;
Gaps	2;			
QY	13	ACTGCATTCCGAAGAGCTAGAGGAAAGTTGCCCTTATACAGAGGAGCCAGTGA	72	
Db	153	ACTCCATCCCAAGAGGCTGGAAGGGAGTTGTCACCGCGGGCGAGGGG	212	
QY	73	ATTGGAGAAACCAACAGCAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTGCTGAT	132	

```

; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1288 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
; US-08-440-856A-9

Query Match      12.2%; Score 100; DB 1; Length 1288;
Best Local Similarity 47.9%; Pred. No. 4.3e-22;
Matches 292; Conservative 39; Mismatches 259; Indels 20; Gaps 4;

Db 213 ATCGGGAGCGCATGTGAGGCTGTTCGTTAAGCACGGGGCCAAAGGTGGTGATCGCGGAC 272
Qy 133 GTCCAAGATGAATAGTTCACCTCAGTTCGTCAGGCGCATTTGGCACTTCCAATTCACCTAC 192
Db 273 ATCGACGACGCGGGCGGCGGCGCTGGCGGGCGGCTGGG--GCGCGACGTGCGGTTTC 329
Qy 193 ATCCACGTGATGTACTAATGAAGACGGTGTAAATATGCGTGGACACACAGTTTCA 252
Db 330 GTGCGGTGCGGCGTGTGCGTGGAGGAGGACGTGGAGCGCGCGTGGAGCGCGCGTGGCG 399
Qy 253 ACCTATGGAATGACATATGTTAGCAATGTCAGCAATGTCAGGAAATTTCTGATCCCAACAGCC 312
Db 390 CGGTACGCGGCGGTGACGTGTGTGCAACACGCGCGGGGTGCTGGCGCCGCCAGACGCGC 449
Qy 313 -----CGCATATAGACAAACAAAAGCAGACATTTGAACGCGTTTCTCAGTGTAAAT 363
Db 450 GCCGCCAAGAGCATCTGCTGCTCGACGCGGGGAGTTTCGACCGCGTCTCCGCGTCAAC 509
Qy 364 GTACCCGAGCTTTTCCTATGTCATGAAGCAGCAGCAGCTGTATGATTTCCAGCAGCGAGT 423
Db 510 GCGTGGCGCGCGCTCGGCATGAAGCAGCGCGGCGCTCGCCATGACCCAGCGCGCGCC 569
Qy 424 GGCAACATATTTCCACTCTCTAGTTTAAGCTCAACTATGCGTGTGCTTCTTCAATGCC 483
Db 570 GCGAGCATATCTCGCTCGCAGCGCTCGCGCGGCGTCTCGCGCGCTCGCGCGCGCGCC 629
Qy 484 TATTGTTGTTCAAGCATCTGTGTAGCCCTTACTAGGAATCTGGCACTCGAGCTCGGA 543
Db 630 TACACCCCTTCAAGCAGCCCATCTGTGGGCTCACCAGAAGCCGCGCTCGAGCTCGGC 689
Qy 544 CAATTGGCAATAGGTTTAATTGTTGCTCTTTTCGGGCTTCTTACGGCTTTAGGCAAG 603
Db 690 GCCACGGCATCGCGTCAACTGCATCTCCCGCTTCGCGCGTGGCGCCACCGGATGCTCATC 749
Qy 604 AA 605
Db 750 AA 751

RESULT 2
US-08-440-856A-9
; Sequence 9, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELLAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
;   TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE. N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,856A
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLMAN, ROBERT A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 05463-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1517
; TELEFAX: (202) 887-0763
; TELEX: 706141
```

```

; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1288 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
; US-08-440-856A-9

Query Match      12.2%; Score 100; DB 1; Length 1288;
Best Local Similarity 47.9%; Pred. No. 4.3e-22;
Matches 292; Conservative 39; Mismatches 259; Indels 20; Gaps 4;

Qy 13 ACTGCATTCGAAGAAGGCTAGAAGAAAAGTTGCCCTTATAACAGGAGGAGCGACAGTGA 72
Db 231 ACWCASATGCCAAGAGGCTGGAMGGAAAGTGGCCATTGTSAAGCGGCGGCGGCGGG 290
Qy 73 ATTGGAGAAACACACAAAACCTTCTCCCAACATGGAGGCAAAAGTTGCCATTGCTGAT 132
Db 291 ATCGSGAGGCSATCTGTMGGCTGTTTCGYAAGCACGGGCGCMRGGTGTTGATCGCGAC 350
Qy 133 GTCCAAGATGAATTAAGTTCACATGTTGTCAGGAGCCATTGGCACTTCCAATTCACCTAC 192
Db 351 ATCGACGACCCGCGGSGAGGCGCTGGCGKCGCGCTGGGS---CCGASGTCRGS TTC 407
Qy 193 ATCCACTGTGATGTACTAATGAAGACGGTGTAAATATGCGTGGAGACACACAGTTTCA 252
Db 408 GTGCGSTGCCACGCTGCTSGTGGAGGAGACGTSRCCGCGCGTSGAGTGGCGCGTSG 467
Qy 253 ACCTAT---GGAAAATGGACATATATGTTACAAATGACAGGAATTTCTGATCCCAACAG 309
Db 468 CGSRYGRCGCGCGCTSGACGTSYSTGCAACAAACGCGGGGTGCTGGGCGCGCACAG 527
Qy 310 CCC-----CGCATATAGACAAACGAAAAGCAGACTTTGAACGCGTTCTCAGTGTA 360
Db 528 CGCGCGGCCARGAGCATCTCTTCGACGCGSSGSGAGTTTCGACCGCGTCTCGCGTC 587
Qy 361 AATGTAACCGAGCTTTTTCCTATGTCATGAAGCAGCAGCAGCTGTATGA-----TTCAG 415
Db 588 AAGCGCTGGCGCGCGCTCGGSATGAAGCACGCGGCGCKCGCATGACCCAGCGCGC 647
Qy 416 CAGCAGTGGCAACATAATTTCACCTGCTAGTTTAAGCTCAACTATGGTGTGTTCTT 475
Db 648 GCCGCGGGGAGCATCTCTCGTCGCCAGCGTGCSSGCGTCTSGCGCGCTCGGCC 707
Qy 476 CACATGCTATTGTTGTTCAAGCATGCTGTAGCCCTTACTAGGAATCTGCAGTCG 535
Db 708 CGACGCTACACCGCTCCAAGCAGCCATCGTSGGCTCACCAGAAGCCGCTGCG 767
Qy 536 AGCTCGGCAATTTGGCATTAGGTTAATTTGTTCTCTCTTCGCGGCTTCCTACGGCTT 595
Db 768 AGCTSSGCGSCACGGSRTCCGSGTCAACTGCTCTCCCTTCGCGGTGCCACSCCSA 827
Qy 596 TAGGCAAGAA 605
Db 828 TGCTCATCAA 837

RESULT 3
US-08-440-856A-1
; Sequence 1, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELLAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
;   TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE. N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
```

ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release 1.0, Version 1.25
CURRENT APPLICATION DATA:
FILING DATE: 15-MAY-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: WILLMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 05463-20001.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1517
TELEFAX: (202) 887-0763
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1236 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-440-856A-1

Query Match 11.9%; Score 97.2; DB 1; Length 1236;
Best Local Similarity 51.1%; Pred. No. 3.3e-21;
Matches 312; Conservative 0; Mismatches 283; Indels 15; Gaps 3;

QY 8 TTCGAACCTCATTCCGACGAAGCTTAGAGGAAAGTTGCCCTTTATACAGGAGGACCA 67
DB 220 TGCCACACCGGATGCCAAGAGGCTGGACGGGAAGGTGGCCATTGTGACGGCGGCGGC 279
QY 68 GTGGAAATCGAGAAACACAGCAAACTCTCTCCCAACATGAGCCAAAGTTGCCATTG 127
DB 280 GCGGATCGGAGGCGCATCTGCGGCTGTCGCCAAGCAGCGGGCGCGGGTGGTGTGCG 339
QY 128 CTGATGTCCAAAGTAATTAAGTCACTAGTGTCTGAGGCCATTGGCAATTCCTCA 187
DB 340 CGGACATCGACGACGCGCGGGGAGGCGCTGCGTFCGCGCTGGGCCGC ---AGGTCA 396
QY 188 CCTACATCAGTGTATGTTACTAATGAAGAGCGGTGTTAAATGCGCGTGGACACACAG 247
DB 397 GCTTCGTGCGCTGCGACGCTGTCGTGGAGAGGACGTCGCGGCGCGCGTGGACTGG 456
QY 248 TTTCAACCTA---TGGAAACTGGACATTATCTTCAGCAATGCAGAAATTTCTGATCCA 304
DB 457 TGTGCGGCCACGCGCGCGCGCTCGAGCTTACTGCAACCAACCGCGGGGTGCTGGCG 516
QY 305 ACAGGCC-----CGCATCATAGACAAGCAAGCAAAAGCAGACTTTGAACGCTTCTCA 355
DB 517 AGACGCGCGCGCCAGGACATCTCTCTTCGACGCGCGCGAGTTCGACGCGGTCTCC 576
QY 356 GTGTAATGTACCGGAGTTTTCCTATGATGATGAGCAGCAGCAGCGTGTATGATTCAG 415
DB 577 GGTCAACGCGTGGCGCGCGCGCTCGGATGAAGCAGCAGCGCGCGCCATGGCGCGCG 636
QY 416 CAGCAGTGGCAACATAATTTCCACTGCTAGTTAAGCTCAACTATGGTGGTGGTCTTT 475
DB 637 GCGCGCGGGGAGCATCTCTCGTCCGCGCAGCGTCCGCGCGGTGCTGGCGCGCTCG 696
QY 476 CACATGCTATTGTGGTTCAAGAGCATGCTGTGTAGCCCTTACTAGGAATCTGGCAGTGC 535
DB 697 CGCAGCCTACACCGCTCCAGACGCGCATCTGTCGGGTCTACCAAGACGCGCGCTGCG 756
QY 536 AGCTCGGACAATTTGGCAATTAGGCTTAATTTGTTGCTCTCTTTTGGGCTTCTTACG 595
DB 757 AGCTGCGCGCAGCGGGTCCGGGTCAACTGCTCTCGCCCTTCGCGCTGCGCCAGCCCA 816
QY 596 TAGGCAGAA 605
DB 817 TGCTCATCAA 826

RESULT 4

US-09-504-358-13
; Sequence 13, Application US/09504358
; Patent No. 6365376
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; APPLICANT: Brzostowicz, Patricia C.
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIA
; FILE REFERENCE: BC1001 US NA
; CURRENT APPLICATION NUMBER: US/09/504,358
; CURRENT FILING DATE: 2000-02-15
; EARLIER APPLICATION NUMBER: 60/120,702
; EARLIER FILING DATE: 1999-February-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Brevibacterium sp HCU
US-09-504-358-13

Query Match 8.6%; Score 70.2; DB 4; Length 771;

Best Local Similarity 46.2%; Pred. No. 1.1e-12;

Matches 348; Conservative 0; Mismatches 393; Indels 12; Gaps 3;

QY 29 GGCTAGAGGAAAGTTGCCCTTATAACAGAGGAGGAGCCAGTGGAAATGGAGAAACACAG 88
DB 8 GACTCGCGGAAAGTAGCAGTCATTACTTGGGGGCCCGCAGGCATGGGGCGCATACAGT 67
QY 89 CAAAACCTCTTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGTCTCAAGATGAATTAG 148
DB 68 CTGAACCTGTATGCGAGTGGAGGTGCAACAGTAGAGCGTAGTAGATGTCAATGAACAAGA 127
QY 149 GTCACCTCAGTTGTCGAGGCCATTGGCAGCTTCCAA-----TTCCACCTACATCCACTGTG 202
DB 128 GCGCTGCCACTGCCGATGCGATAAGGGCCAGGGCGGGTTCGCAACTATTGGAATTGG 187
QY 203 ATGTTACTAATGAAGCGGTGTTAAATAATGCGTGGACACACAGTTCACACCTATGGAA 262
DB 188 AGTTTCTGACGAGTCTGAAGTTGAAATAGTCGTCTCCGACATTCGCAAGAGATTCGGTG 247
QY 263 AACTGACATTTATGTTACGAATGACAGGAATTTCTGATCCCAACGCGCCCGCATCATAG 322
DB 248 CGATTAACTACTAGTGAACACGAGCGCTCACCGTGCAGTAACTCACTCAC---G 304
QY 323 ACAACGAAAGACGACTTTGAACGCGTCTCAGTGTAAATGTAAACCGAGTTTCTCTAT 382
DB 305 AGATCGACGAACGGGACCTGGACCTCGTACTGAGCGTGTGTAAGAGGATTTCTTCA 364
QY 383 GCATGAAGCACGACGACCGTGTATGATTCAGCAGCAGTGGCAACATAATTTCCACTG 442
DB 365 TGACAAACACATGCTATCCCTTCTTAAACAGCTGGCGGCGGAGCCATCGTCAACTTCG 424
QY 443 CTAGTTTAAAGCTCAACTATGCGTGGTGGTCTTCTACATGCTTATGTTGTTCAAGCATG 502
DB 425 CGTCTATCTATGTTGTTGGGTCCGAGGAGCTTACCCCGTACCACGAGCCAAAGGTG 484
QY 503 CTGTGTTAGCCCTTACTAGGAATCTGGCAGTGCAGCTCGGACAAATTTGGCATTAGGTTA 562
DB 485 CGTCTGTTCCCTTACCAACAGGACGCGGTGACTTACGCGCGCTCAATATCCGAGTGA 544
QY 563 ATTGTTGTCCTTTCGGGCTTCTACGGCTTAGGCAAGAAATTT---TTCAGGAGTTA 619
DB 545 ATGCGGTAGCAGCCCGGAACCATTTTGACTCCACTAGTCAAGGAGCTCGGTTTCAAGGGGC 604
QY 620 AAAATCAAGAAATTTGAGAAATGTAATAAATTTCCGGGAATTTGAAGGTGCCAAAT 679
DB 605 CCGATGGCTTAGATGATATACTAACTTATGGTGCCAGCATCCGCTTGTGCGGTAG 664
QY 680 TTAATGTTGAGGATGTGCCAATGCAAGCTCTTTATCTGGCTAGTAGGCAAAATACG 739

Db 665 GAACCCCGAAGAAGTCGGCGCAGCAACATCTTTCTGGCATCCGAAGAAGCTTCGTTC 724

Qy 740 TGAGTGGACACAATCTGTTTCATTGAGGAGGT 772

Db 725 TTAAGTGGCGCGTCCCTCCGCTTGAGGAGGT 757

RESULT 5

US-09-954-314-13

Sequence 13 Application US/09954314

Patent No. 6465224

GENERAL INFORMATION:

APPLICANT: Rouviere, Pierre E.

APPLICANT: Brzostowicz, Patricia C.

TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES

FILE REFERENCE: BC1001 US NA

CURRENT APPLICATION NUMBER: US/09/954.314

PRIOR FILING DATE: 2001-09-17

PRIOR FILING DATE: 60/120,702

NUMBER OF SEQ ID NOS: 49

SOFTWARE: Microsoft Office 97

SEQ ID NO 13

LENGTH: 771

TYPE: DNA

ORGANISM: Brevibacterium sp HCU

US-09-954-314-13

Query Match 8.6%; Score 70.2; DB 4; Length 771;

Best Local Similarity 46.2%; Pred. No. 1.le-12;

Mismatches 0; Mismatches 393; Indels 12; Gaps 3;

Db 29 GGCTAGAGCAAAAGTGGCCCTTATAACAGGAGGAGCCAGTGGAAATGGAGAACACACAG 88

Db 8 GACTCGCGCAAGAGTAGCATCTATTACGAGGCGCCGCGGAGGATGGGGCGCATACAGT 67

Qy 89 CAAACTCTTCTCCCAACATGAGCCAAAGTTGCCATTGCTGATGTCACAGATGAATAG 148

Db 68 CTGAACGTGATGCGAGTGGAGGTGCACAGTAGCGGTAGTAGATGTCAATGAACAAGA 127

Qy 149 GTCACCTCAGTTCGAGGCGCATGGCAGTTCACAA-----TTCCACCTACATCCACTGTG 202

Db 128 GCCGTGCCACTCGCGATGGGATAGAGGCGCGCGGGGTTCGCAACTATTGGAAATGG 187

Qy 203 ATGTTACTAATGAAGCGGTGTTAAAAATGCGTGAACACACAGATTTCAACCTATGAA 262

Db 188 ACGTTCTGACGAGTCTGAAGTTGAATAGTCTGTCGACATTCGCAAGAGATTCGGTG 247

Qy 263 AACTGGACATTTATGTCAGCAATGAGGAATTTCTGATCCCAACAGCGCCGCGCATCATAG 322

Db 248 CGATTACGTAAGTGAACAACGCGGCGTCCAGCGGTGCGGATTAACCAACTCAC---G 304

Qy 323 ACAACGAAAAGCAGACTTTGAACGCGTCTCAGTGAATGTAACCGGAGTTTCCGTAT 382

Db 305 AGATCGACGACGCGGACCTGGACCTGCTACTGAGCGTCTGTAAGAGGATTTCTTCA 364

Qy 383 GCATGAGCAGCAGCAGCTGTTATGATCCAGCAGCGAGTGGCAACATAATTTCCACTG 442

Db 365 TGACAAACACTCGATCCCTTACTTTAAACAGCGTGGCGGCGGAGCCATCGTCAACTCG 424

Qy 443 CTAGTTTAAAGCTCAACTATGGTGGTGTCTCATGCTCATGTCATGTTGTTTCAAGCATG 502

Db 425 CGTCTATGCTATGCTGGTGGGTGCGAGGAGCTTACCCCGTACCGACGAGCAAGAGGTG 484

Qy 503 CTGTTGTTAGCCCTTACTAGGAATCTGGCAGTGGAGCTCGGCAATTTGGCATTAGGGTTA 562

Db 485 CGGTCTGTCCTTCAACAAACAGGACGCGGTGACTTAGGACCGTCAAAATCCGAGTGA 544

Qy 563 ATGTTGTTGCTCTCCCTTCAGCGGCTTACGGGTTTAGGCAAGAAATTT---TTCAGGATTA 619

Db 545 ATCGGTGAGCACCCGCAACATTTTGAATCCACTAGTCAAGGAGCTCGGTTCAAGGGCC 604

Qy 620 AAAATGAAGAAGATTTGAGATTAATAAATTTTTCGGGGAAATTTGAAGGTTCCAAAT 679

Db 605 CCGATGCTTAGATGGATATACTAACTTATGGTGCCCAAGCATCCGCTTGGTGGGTAG 664

Qy 680 TTAATGTTGAGGATGTTGCCAATGACAGCTTTATCTGCTAGTATGATGAGCAAAATACG 739

Db 665 GAACCCCGAAGAAGTCGGCGCAGCAACATTTGTTTGCATCCGAAGAAGCTTCGTTC 724

Qy 740 TGAGTGGACACAATCTGTTTCATTGATGGAGGT 772

Db 725 TTAAGTGGCGCGTCCCTCCGCTTGAGGAGGT 757

RESULT 6

US-09-504-358-15

Sequence 15 Application US/09504358

Patent No. 6365376

GENERAL INFORMATION:

APPLICANT: Rouviere, Pierre E.

APPLICANT: Brzostowicz, Patricia C.

TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIA

FILE REFERENCE: BC1001 US NA

CURRENT APPLICATION NUMBER: US/09/504.358

PRIOR FILING DATE: 2000-02-15

PRIOR FILING DATE: 60/120,702

NUMBER OF SEQ ID NOS: 49

SOFTWARE: Microsoft Office 97

SEQ ID NO 15

LENGTH: 10629

TYPE: DNA

ORGANISM: Brevibacterium sp HCU

US-09-504-358-15

Query Match 8.6%; Score 70.2; DB 4; Length 10629;

Best Local Similarity 46.2%; Pred. No. 5.le-12;

Mismatches 348; Mismatches 393; Indels 12; Gaps 3;

Qy 29 GGCTAGAGCAAAAGTGGCCCTTATAACAGGAGGAGCCAGTGGAAATGGAGAACACACAG 88

Db 9139 GACTCGCGCAAGAGTAGCATCTATTACGAGGCGCCGCGGAGGATGGGGCGCATACAGT 9198

Qy 89 CAAACTCTTCTCCCAACATGAGCCAAAGTTGCCATTGCTGATGTCACAGATGAATAG 148

Db 9199 CTGAACGTGATGCGAGTGGAGTGCACAGTAGCGGTAGTAGATGTCAATGAACAAGA 9258

Qy 149 GTCACCTCAGTTCGAGGCGCATGGCAGTTCACAA-----TTCCACCTACATCCACTGTG 202

Db 9259 GCCGTGCCACTCGCGATGCGATAAGGCGCGCGGGGTTCGCAACTATTGGAAATGG 9318

Qy 203 ATGTTACTAATGAAGCGGTGTTAAAAATGCGTGAACACACAGATTTCAACCTATGAA 262

Db 9319 ACGTTCTGACGAGTCTGAAGTTGAATAGTCTGTCGACATTCGCAAGAGATTCGGTG 9378

Qy 263 AACTGGACATTTATGTCAGCAATGAGGAATTTCTGATCCCAACAGCGCCGCGCATCATAG 322

Db 9379 CGATTAAAGCTAGTAGTGAACAGCGGCGTCCCGGTGCGGATTAACCAACTCAC---G 9435

Qy 323 ACAACGAAAAGCAGACTTTGAACGCGTTCCTCAGTGAATGTAACCGGAGTTTTCCTAT 382

Db 9436 AGATCGACGACGCGGACCTGGACCTGCTACTGAGCGTGTGATGTAAGAGGATTTCTTCA 9495

Qy 383 GCATGAGCAGCAGCAGCTGTTATGATCCAGCAGCGAGTGGCAACATAATTTCCACTG 442

Db 9496 TGACAAACACTCGATCCCTTACTTTAAACAGGCTGGCGGCGGAGCCATCGTCAACTCG 9555

Qy 443 CTAGTTTAAAGCTCAACTATGGTGGTGTCTTCATGCTCATGTCCTATTTGTTTCAAGCATG 502

Db 9556 CGTCTATCTATGCTGGTGGGTGCGCAGGAGCTTACCCCGTACCGACGCAAGAGGTG 9615

Qy 503 CTGTTGTTAGCCCTTACTAGGAATCTGGCAGTGGAGCTCGGCAATTTGGCATTAGGGTTA 562

Db 9616 CGGTGCTGCTCTTACCAACAGGAGCGCGTCACTTACGGCCGTCAAAATATCCGAGTGA 9675

Db 247 GATGTTTGTGACAAATGACAGGAT-----TACCCAGATACCTTATGCTCAAGATG 300
Qy 328 GAAAAACAGACTTTGAACGCGTTCFAGTGTAAATGTAAACCGAGTTTTCCTATGATG 387
Db 301 ACAGAACAGATTTTGAAGAGTGTCTCAAGGTCAATCTGACTGTGCTTTAAATATGACA 360
Qy 388 AAGCAGCAGCAGCTGTATGATTCAGCAGCAGTGGCAACATAATTTCCACCTGCTAGT 447
Db 361 CAATCAGCTTTGAACCGATGATGAAGCCAGAGAGGTGCTATCATTAATATGCTAGT 420
Qy 448 TTAAGCTCAACTATGGTGTGTTCTTCACATGCTATGTTGGTTCAAGCATGCTGTG 507
Db 421 GTTCTGTTGATGGGATATGCTCAAGCTAACTATGCTGCTTCAAGGCTGGCTTG 480
Qy 508 TTAGCCCTTACTAGGATATGCGAGTTCGAGCTCGACACTTGGCAATTAGGGTTAATGT 567
Db 481 ATTGGCTTTACCAAGTCTGTGGCCGCGAGTGTGCTAGTCGGAATATACGAGTCAATGTG 540
Qy 568 TTGTCCTCTTTCGGGCTTCTAGCGGCTTTAGCGCAAGAAATTTTCAGGGATTAAAAATGAA 627
Db 541 ATTGCTCCAGGAATGATGAGTCTGATATGACAGCTATCTTATCAGATAAGATTAAAGAA 600
Qy 628 G 628
Db 601 G 601

RESULT 9
US-08-961-527-7/c
; Sequence 7, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-7

Query Match 7.3%; Score 59.6; DB 4; Length 19702;
Best Local Similarity 46.1%; Pred No. 1.8e-08;
Matches 280; Conservative 0; Mismatches 319; Indels 9; Gaps 2;

Qy 24 AAGAGGCTAGAGGAAAGTTGCCCTTTATAACAGAGGAGGCCAGTGAATTTGGAGAAC 83
Db 16196 AATGAAGTAAAGATAAAATATCTTTATTACAGGTTCGAGTCGTGAATTTGTTCTTGC 16137
Qy 84 CACAGCAAACTCTTCTCCCAACATGGAGCAAAAGTTGCCATTTGCTATGATCTCCAGATGA 143
Db 16136 CATGCCCAACAAGTTTGTCTCAAGCAGGAGCAACATTTGCTTTAAACAGTCTGTTGGGCAAT 16077
Qy 144 ATTAGGTCAGTCACTGTTGTCGAG--GCCATTGGCACTTCCCAATTTCCACCTACATCCACTG 200
Db 16076 CTCAGAAGAATTTGCTGCTGAGTTTTCAAACATATGTTCAAGGTGTTTCCCATTTTCAGG 16017
Qy 201 TGATCTTACTAATAAGACAGCGTGTAAAAATGCCGTGGACACACAGTTTCAACCTATGG 260
Db 16016 AGATGATCAGATTTTTCAGACCGCTAAGCGTATGATTTGATCAAGCTATTTGCAGAACTGGG 15957
Qy 261 AAAACTGGACATTTATGTTTCAGCAATCGAGAAATTTCTGATCCCAACAGGCCCGCATCAT 320
Db 15956 TTCAGTATGATGTTTGGTCAACAATGCAGGGAT-----TACCCCAAGATACTCTTTATGCT 15903
Qy 321 AGACAACGAAAAAGCAGACTTTGAACGCGTTCTCAGTGTAAATGTAAACGGAGTTTTCCT 380
Db 15902 CAAGATGACAGAAGCAGATTTTGAAGAAAGTGTCTCAAGGTCAATCTGACTGGTCTTTAA 15843
Qy 381 ATGATGAAGCAGCAGCAGCTGTTATGATTTCCAGCAGCAGTGGCAACATAATTTCCAC 440
Db 15842 TATGACACAATCAGTCTTGAAACCGATGATGAAGCCAGAGAAGTGTCTATCATTAATAT 15783
Qy 441 TGCTAGTTTAAAGCTCAACTATGTTGGTGTGTTCTTTCACATGCTATTGTTGTTCAAGCA 500
Db 15782 GTCTAGTGTGTTGGTTGATGGGAATATTTGTTCAAGCTAATTTATGCTGCTTCTAAGGC 15723
Qy 501 TGCTGCTTACGCTTACTAGTAATCTGCGAGTTCGAGTTCGAGCAAAATTTGGCATTAGGT 560
Db 15722 TGGCTTATTTGCTTTTACCAAGTCTGTGGCACCGAGTGTCTAGTCGGAATATACAGT 15663
Qy 561 TAATGTTTGTCTCTCTTTCGGGCTTCTCAGCGCTTCTAGCGCTTTAGGCAAGAAATTTTCAGGATATA 620
Db 15662 CAATGATGTTGCTCCAGGAATGATGAGTCTGATATGACCGCTATCTCTGCAGATAAGAT 15603
Qy 621 AAATGAAG 628
Db 15602 TAAGGAAG 15595

RESULT 10
US-08-858-207A-8
; Sequence 8, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328e1 Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997

```

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1502
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1502

Query Match 6.6%; Score 54.4; DB 4; Length 873;
Best Local Similarity 62.5%; Pred. No. 1.4e-07;
Matches 85; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 22 GCRAAGAGCGCTAGAGAAAGTTGCCCTTATACAGGAGGAGCCAGCTGGAATTGGAGAA 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 GCAGGAAAGTTAGAGATAAAGTTGCTTAAACAGGAGGAGATTCCAGGTATTGGACGT 177

QY 82 ACCACAGCAAACTCTTCTCCACACTGGAGCCAAAGTTGCCATTGCTGATGTCCAAGAT 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 GCGATAGCAATCTATATGCTAAAGAAGGGGCAATGTTGCTATTGTTATTATGACGAA 237

QY 142 GAATTAGGTCACCTCAG 157
    | || | || | || |
Db 238 CATCAGATCCGAG 253

RESULT 12
US-09-134-001C-725
; Sequence 725, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 725
; LENGTH: 774
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-725

Query Match 6.6%; Score 54; DB 4; Length 774;
Best Local Similarity 49.7%; Pred. No. 1.8e-07;
Matches 172; Conservative 0; Mismatches 165; Indels 9; Gaps 1;

QY 32 TAGAAGGAAAGTTGCCCTTATACAGGAGGAGCCAGTGGAAATGGGAAACCACAGCAA 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35 TAACTGGTAAGGTGGCAATGGTTCATTGGAGGTGCAACTGGCATTTGGTAAAGCTATGGCAG 94

QY 92 AACTCTTTCTCCCAAGTAGGAGCCAAAGTTGCCATTGCTGATGTCCAAGATGAATTAGGTC 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 AGCATTAGCAAGCAGAGGGGCAAAATATCGTTATCGCTGATTTGGCAATCCAAATATTGGTC 154

QY 152 ACTCAGTTGTGAGGGCAATTGGCAC-----TTCCAATTCCACCTACATCCACTGTG 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 AAGAAACAGCCACAACATATATCTACACAATCAGTGTGTTAAACCACACATCATTTAAATTAG 214

QY 203 ATGTTACTAATGAGACGGTGTGTTAAAAATGCCGTGGGACACACAGTTTCAACCTATGGAA 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 ATATAACTCATCTAGTAGAAGTAATAATCAAAATGTTGTGATTATGATGAAGGCAATATGCAA 274

```


APPLICANT: Mitsuhashi, Kazuya
TITLE OF INVENTION: NOVEL CARBONYL REDUCTASE, METHOD FOR PRODUCING SAID ENZYME, DNA
TITLE OF INVENTION: ENCODING SAID ENZYME, AND METHOD FOR PRODUCING ALCOHOL USING SAID
FILE REFERENCE: 06501-050001
CURRENT APPLICATION NUMBER: US/09/468,738A
CURRENT FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: JP 1999-171160
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: JP 1998-363130
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0, reformatted using WordPerfect 5.1
SEQ ID NO 28
LENGTH: 786
TYPE: DNA
ORGANISM: Bacillus subtilis
US-09-468-738A-28

Query Match 6.2%; Score 50.4; DB 4; Length 786;
Best Local Similarity 51.8%; Pred. No. 2.5e-06;
Matches 145; Conservative 0; Mismatches 126; Indels 9; Gaps 1;
QY 32 TAGAAGAAAGTTGCCCTTATACAGGAGGAGGCCAGTGGAAATGGAGAAACACAGCAA 91
DB 14 TAAAGGAAAGTCTCGCTATTACAGGAGCTGCTTCAGGGCTCGGAAAGCGGATGGCCA 73
QY 92 AACTCTTCTCCCAACATGGAGCCAAAGTTGCCAT-----TGCTGATGTCCAAGATG 142
DB 74 TTCGCTTCGGCAAGGAGGAGCGCAAAAGTGTATTACACTATTATAGTAATAACAAGATC 133
QY 143 AATTAGTCACTCAGTGTGTCGAGGCCATTGGCATTCCCAATTCCACCTACATCCACTGTG 202
DB 134 CGAACGAGGTAAAGAAGAGGTTCATCAAGCGGGCGGTGAAGCTGTTGTCGTCCAAGGAG 193
QY 203 ATGTTACTATGAAGAGGTGTTAAATAATGCCCTGGACACACAGTTTCAACCTATGGAA 262
DB 194 ATGTCACGAAAGAGGAGATGTAAAAATATCGTGAACCGCAATTAGGAGTTCCGCA 253
QY 263 AACTGGACATTATGTCAGCAATGCAGGAATTTCTGATCC 302
DB 254 CACTCGATTATTATGATTAATATGCCGGTCTTGAAATCC 293

Search completed: February 18, 2003, 17:47:38
Job time : 114 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 17:21:59 ; Search time 114 Seconds
(without alignments)
3659.088 Million cell updates/sec

Title: US-09-673-918A-1

Perfect score: 819

Sequence: 1 atgcagcttcgaactgattc.....tgttccaatatccagattct 819

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	736	89.9	4505	9	US-09-944-160-23
2	150.2	18.3	774	9	US-09-938-842A-945
3	149	18.2	833	10	US-09-770-445-686
4	123.6	15.1	391	10	US-09-878-574-2791
5	117.8	14.4	358	10	US-09-878-574-98
6	87.4	10.7	375	10	US-09-878-574-3102
7	78	9.5	6948	10	US-09-070-927A-4
8	70.2	8.6	771	10	US-09-954-314-13
9	70.2	8.6	10629	10	US-09-954-314-15
10	62.4	7.6	774	9	US-09-891-641-55
11	60.6	7.4	732	10	US-09-815-242-9262
12	60.6	7.4	732	10	US-09-815-242-9483
13	60.6	7.4	762	10	US-09-815-242-7745
14	56.8	6.9	805	10	US-09-770-445-797
15	54.8	6.7	762	10	US-09-974-300-2259
16	52.8	6.4	2774	10	US-09-802-853-5
17	51	6.2	735	10	US-09-815-242-6029
18	50.4	6.2	786	10	US-09-940-037A-28
19	49.2	6.0	1716	10	US-09-796-089-10

ALIGNMENTS

RESULT 1

US-09-944-160-23
; Sequence 23, Application US/09944160
; Patent No. US20020174452A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, No. US20020174452A1man
; APPLICANT: Davin, Laurence
; APPLICANT: Huang, Ning
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan
; TITLE OF INVENTION: Content
; FILE REFERENCE: WSUR17983
; CURRENT APPLICATION NUMBER: US/09/944,160
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 60/230,632
; PRIOR FILING DATE: 2000-09-07
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 4505
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid pAPI249 encoding secoisolariciresinol
US-09-944-160-23

Query Match 89.9%; Score 736; DB 9; Length 4505;
Best Local Similarity 93.9%; Pred. No. 4.2e-204;
Matches 766; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY	4	CAGCTTCGAACTGCATTCCGAAGAGCTAGAGGAAAAGTTCCCTTATAACAGGAGGA	63
DB	965	CAGCTTCGAACTGCATTCCGAAGAGCTAGAGGAAAAGTTCCCTTATAACAGGAGGA	1024
QY	64	GCCAGTGGAAATTGGAGAAACACAGCAAAACTCTTCTCCCAACATGGAGCAAAGTTGCC	123
DB	1025	GCCAGTGGAGTTGGAGAAAGTCACAGCAAAACTCTTCTCCCAACATGGAGCAAAGTTGCC	1084
QY	124	ATTGCTGATGTCCAAAGATGAATAGTCTACTGATGTGTGAGGCCATTGGCACTTCCAAT	183
DB	1085	ATTGCTGATGTCCAAAGATGAATAGTCTACTGATGTGTGAGGCCATTGGCACTTCCAAT	1144

APPLICANT: Garcia, Carlos A.
APPLICANT: Krieker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2023US (PARA-012PRV)
CURRENT APPLICATION NUMBER: US/09/770,445
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/178,472
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 686
LENGTH: 833
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(833)
OTHER INFORMATION: n = A,T,C or G

US-09-770-445-686

Query Match 18.2%; Score 149; DB 10; Length 833;
Best Local Similarity 52.2%; Pred. No. 1.4e-33;
Matches 395; Conservative 0; Mismatches 350; Indels 12; Gaps 3;

QY 32 TAGAGGAAAGTTGCCCTTATACAGGAGGAGCAGTGGAAATGGAGAAACACACGAA 91
DB 833 TGGATGGCAAAATCGCAATAATAACAGCGGAGTAGCGGGATTGGAGCGGCGTTA 774
QY 92 AACTCTCTCCCAACATGAGCAGCAAGTTGCCATTGCTGATCTCCAGATGAATAGGTC 151
DB 773 GCGTGTACCGGACACGAGGTAGGTGGTTCATGCTTCAAGAAAGAACTGGTGC 714
QY 152 ACTCAGTGTTCGAGGCGCAATGGCACTTCCAAATTCACCTACATCCACTGTGATCTACTA 211
DB 713 AAAAGCTTCCCGTCTCTGTCGGGAAAGACAAAGCANNNTTTACCGTTCGATGTACAA 654
QY 212 ATGACAGCGGTGTTAAATGCGGTGGACACACAGTTTCAACCTATGAAACATGGACA 271
DB 653 ACGAAAGAGGTTNNAACGCGGTTAAGTTCACCGTCGAAAGTACGGGAAGCTTGAGC 594
QY 272 TTATGTTACAGCAATGCAGGAATTTCTGATCCCAACAGCGCCGCATCATAGACACGAA 331
DB 593 TTCTCTTTNNACCGCGCGTATGGAAC---AGCCGGGAAGCTTCTCGACTTGAATC 537
QY 332 AAGCAGACTTTGAACGCGTCTCTCAGTGAATGTAAACCGGAGTTTTCCTATGCATGAAGC 391
DB 536 TGGACAGTTTACCGCAACCATGGCGTCAACGTTCTGTGCGGCTGCGTTATCAAC 477
QY 392 ACGAGCAGCTTTATGATTCAGCAGCA---GTGGCAACATAATTTCCACTGTAGTT 448
DB 476 ACGCTGCAGCGCATGTGGAGAAAGGACGCGGTGCTCAATCGTATGTACGACGAGG 417
QY 449 TAAGCTCAACTATGGTGGTGTCTTCCATGCTCAATGTTGTTCAAGCATGCTGCT 508
DB 416 TCGGTGCGAGATCGGTGTCCAGGACCTACGCGGTACACGCGGTCTAAGCAGCTCTTC 357
QY 509 TAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAAATTTGGCATTAAGGTTAATGTT 568
DB 356 TCGGCTGTTTAAATCGGCTTGTGGCGGCTAGGAGTACGGGATAGAGTCAACGCGG 297
QY 569 TGTCTCTTTCGGGCTTCTAGCGCTTTAGGCAAGAAATTTTCAGGATTAATAATGAAG 628
DB 296 TTGCACCATACGCGGTGGCGGCGGATA-----AATAGCGGTGACGAGGAACGGTGA 243
QY 629 AAGAAATTCAGAAATTAATAAATTTGCGGGAATTTGAAAGGTCCAAATTTAATGTTG 688
DB 242 GGATGGTGGAGGAATATAGCGCAGCCAGCGGGGATTTCTCAAGAGGTGTGGTCTTAAGGCTC 183

QY 689 AGGATGTTGCCAATGCAGCTCTTTATCTGGCTAGTATGAGGCAAAATACGTAGTGGAC 748
DB 182 GCCATGTGGCAGAGCGGCTTTTGTCTGGCTTCGGATGATTCGGCTTACGTAGCGGTC 123
QY 749 ACAATCTGTTTATGATGAGGTTTCAGGCTCTGCA 785
DB 122 AGAATCTGGCTGTTCACGGGGGTTATAGCGTCTTAA 86

RESULT 4

US-09-878-574-2791
Sequence 2791, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrium, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 2791
LENGTH: 391
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)-(391)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: LIB3028-020-Q1-B1-D11

US-09-878-574-2791

Query Match 15.1%; Score 123.6; DB 10; Length 391;
Best Local Similarity 60.8%; Pred. No. 2.3e-26;
Matches 226; Conservative 0; Mismatches 134; Indels 12; Gaps 1;

QY 432 AATTCCACTGCTAGTTTAAGCTCAACTATGGTGGTGTCTTTCACATGCCTATTTGTGG 491
DB 1 AATTAACACTGCTAGTTTGTCTGGAACCTTTAGTGGAGGGGCTTCACATGCTACACAAG 60
QY 492 TTCAAGCATGCTGTAGCCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGG 551
DB 61 TTCAAGCAGCAGCACTAATTTGGACTGATGAAACACTGCGGTGGAGCTTGGACAGTTGG 120
QY 552 CATTAGGGTTAATTTGTTCTCTCTTCGGGCTTCTACGGCTTTTAGGCAAGAAATTTTC 611
DB 121 TATTAGGGTTAATTTGTTCTCTCTCTCTATGTTGCCACACCATTCACCTAAGAAATGTTT 180
QY 612 AGGGATTAATAATGAAGAATTTGAGAAATGTAATAAACTTTGGGGGAAATTTGAAGG 671
DB 181 CAATCTTGTATGAAGACCGAAATGGTGA-----TTTATTCACACCTAAAGG 228
QY 672 TCCAAATTAATTTGAGGATGTTGCCAATGCAGCTCTTTATCTGCTAGTGCATGAGGC 731
DB 229 TGTTCATCTGTGCCAAGCATGTGGCGGAAGCTGCTCTATATTTGGCAGGTGATGAGTC 288
QY 732 AAAATACGTGAGTGGACACAATCTGTTCAATTTGATGGAGGTTTCAGCGTCTGCAATTTCTGT 791
DB 289 AAAGTATGTTAGTGGTCACAATCTTCTGTAGATGGAGGTTTCACCAATCTAAATCTAGG 348
QY 792 AATCAAGTGT 803
DB 349 ATTTCTGTGT 360

RESULT 5

US-09-878-574-98
Sequence 98, Application US/09878574

Query Match 9.5%; Score 78; DB 10; Length 6948; Best Local Similarity 47.2%; Pred. No. 2e-12; Matches 274; Conservative 0; Mismatches 300; Indels 6; Gaps 1;									
QY	32	TAGAAGAAAGTTGCCCTATATACAGGAGGAGCCAGTGGAAATGGAGAAACCCACAGCAA	91						
DB	3145	TATCAGAAAAGTTATGTGTATCATGGCGCTTCCAGTGGCATCGGTGAGGCAACAGCCC	3086						
QY	92	AACCTTCTCCCAACATGGAGCCAAAGTTGCCATGTGATGCCAAGATGAATTAGGTC	151						
DB	3085	GTTTACTTGCAGAAAGGAGCGAAGTAGTCATTGCAGCAGCTGCACRAGAAGCTTAA	3026						
QY	152	ACTCAGTTGTGAGGCCATTTGGCAGCTTCCAAATTCACCTACATCCACTGTGATGTTACTA	211						
DB	3025	TTGCTATAAAAAAGAACTTCTTGAAGCACAGATTCTCGTGCACACAGAGATGTAACGA	2966						
QY	212	ATCAAGACGGTGTAAAAATGCCGTGGACAACACAGCTTTCAACCTATGGAAAACCTGCACA	271						
DB	2965	AAGAAGAGGAAAGTTCAGGTGTAAATCAAGCTTACAATGGAAAATATGGAGAAATTTGATG	2906						
QY	272	TTATGTTCCAGCAATGCAGGAATTTCTGATCCCAACAGGCCGCCATCATATAGACAAGGAA	331						
DB	2905	TTCTTTTAAACAATGCAGGA-----GTTATGCCAACGGCGCTCTTATTGAGACCCAA	2852						
QY	332	AACACAGCTTTGAACGGTTCCTCAGTGTAAATGTAACCGGAGTTTCCCTATGATGATGAAGC	391						
DB	2851	AAGGAAATGGCGTCAATGCTAGATATTAATATTATGGGTGTTTAAATGGCATTTCCGG	2792						
QY	392	ACGACAGCAGTGTATGATTCACAGCAGCAGTGCGCAACATAATTTCCACTGCTAGTTTAA	451						
DB	2791	CAGTTCTTCCAATATGTTTGAACAAAATCAGGGCAGATATTGCTACTGATCTCTGTGG	2732						
QY	452	GCTCAACTATGGTGGTGTCTTCACATGCCCTATTGTGTTCAAAAGCATGCTGTGTAG	511						
DB	2731	CAGGACATGTCGTTATCCAGATTGCGCTCTACTGTGGAACAAGTTGCTGTTCGAG	2672						
QY	512	CCCTTACTAGGAATCTGGCAGTCGAGCTCGGACAATTTGGCATTAGGGTTAATGTTGT	571						
DB	2671	CAATTATGGAAGGATTCGCAACAAGAACAAAGAGAGAATAACATTAATCAACGATTATT	2612						
QY	572	CTCCTTTCCGGCTTCTACGGCTTTAGGCAAGAAATTTTC	611						
DB	2611	CACCAGTGCTGTACAACAGAACTTTATCAACAATTTTC	2572						
RESULT 8									
US-09-954-314-13									
; Sequence 13, Application US/09954314									
; Patent No. US20020127666A1									
; GENERAL INFORMATION:									
; APPLICANT: Rouviere, Pierre E.									
; APPLICANT: Brzostowicz, Patricia C.									
; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES									
; FILE REFERENCE: BC1001 US NA									
; CURRENT APPLICATION NUMBER: US/09/954,314									
; CURRENT FILING DATE: 2001-09-17									
; PRIOR APPLICATION NUMBER: 60/120,702									
; PRIOR FILING DATE: 1999-February-19									
; NUMBER OF SEQ ID NOS: 49									
; SOFTWARE: Microsoft Office 97									
; SEQ ID NO 13									
; LENGTH: 771									
; TYPE: DNA									
; ORGANISM: Brevibacterium sp HCU									
US-09-954-314-13									
Query Match 8.6%; Score 70.2; DB 10; Length 771;									
Best Local Similarity 46.2%; Pred. No. 1.2e-10;									
Matches 348; Conservative 0; Mismatches 393; Indels 12; Gaps 3;									
QY	29	GGCTAGAGAAAGTTGCCCTATATACAGGAGGAGCCAGTGGAAATGGAGAAACCCACAG	88						
DB	8	GACTCGCGGAAAAGTAGCAGTCATTACTTGGGGCGCCGAGGCATGGGGCCGATACAGT	67						

QY	89	CAAAACTCTTCTCCCAACATGGAGCCAAAGTTGCCATTTGCTGATGCTCAAGATGAATTAAG	148						
DB	68	CTGAACCTGTATGCGAGTGGAGGTGCAACAAGTAGCGGTAGTAGATGTCATGAACAAGAAG	127						
QY	149	GTCACTCAGTTGTGCGAGGCCAATTTGGCAGCTTCCAA-----TTCCACCTACATCCACTGTG	202						
DB	128	GCGTGCCTACTGCGCATGGCATTAAGGCCAGCGCGGGTTGCAAACTATTTGAAAATTGG	187						
QY	203	ATGTTTACTATGAAGACGGTGTAAANAATGCCGTGGACACACAGTTTCAACCTATGGAA	262						
DB	188	ACGTTTCTGACGAGTCTGAAGTTGAATAATAGTCGTCCTCCGACATTTGCCAAGAGATTCCGGTG	247						
QY	263	AACCTGACATTTATTTTCAGCAATGCAGGAATTTCTGATCCCAACAGGCCCGCCGATCATAG	322						
DB	248	CGATTAAACGTACTAGTGAACAACGACGAGCGCTCACCGGTGCCGATAAACCAACTCAC---G	304						
QY	323	ACAACGAAGAAGCAGACTTTGAACGGTTCCTCAGTGTAAATGTAACCGGAGTTTTCCTAT	382						
DB	305	AGATCAGCAAAACGGGACCTGGACCTGCTACTGAGCGTCGATGTGAAGGAGTATTTCTTCA	364						
QY	383	GCATGAAGCACGACGACGCTGTTATGATTCACGACGAGTGGCAACATAATTTCCACTG	442						
DB	365	TGACAAAACACTGCATCCCTACTTTAAACAGGCTGGCGCGGAGCCATCGTCAACTTCG	424						
QY	443	CTAGTTTAAAGCTCAACTATGGTGGTGGTTCCTTTCACATGCTCTATTGTGGTTCAAAGCATG	502						
DB	425	CGTCTATCTATGCTGCTGGTGGGTGCGCAGGAGCTTACCCCGTACCAGCGCAAGAGTG	484						
QY	503	CTGTGTTAGCCCTTACTAGAACTCTGGCAGTCGAGCTCGGACAATTTGGCATTAGGGTTA	562						
DB	485	CGGTCTGTTGCCCTTACCAACAGAGCGCGGTGACTTTACGGACCGTCAAAATATCCGAGTGA	544						
QY	563	ATGTTTGTCTCTCTTTCGGGCTTCTCAGCGCTTTAGSCAAGAATTT---TTCAGGGATTA	619						
DB	545	ATCGGTAGCACCGGAAACCATTTTGACTCCACTGATCAAGGAGCTCGGTTCAAGGGGCC	604						
QY	620	AAATGAAGAAGAATTTGAGAATGTAAATAACTTTGCGGGAATTTGAAAGTCCAAAT	679						
DB	605	CCGATGCTTAGATGGATATATACTAACTTATGGTGCCCAAGCATCCGCTTGGTCCGGTAG	664						
QY	680	TAAATGTTGAGGATGTTGCCAATGCAGCTCTTTATCTGCGTAGTGTAGGCAAAATACG	739						
DB	665	GAAACCCCGAAGAAGTCGCGGCGAGCAACATTTGTTCTGGCATCCGGAAGAGCTTCGTTC	724						
QY	740	TGAGTGGACACAATCTGTTCAATTGATGGAGGT	772						
DB	725	TTACTCGCGCTCTCTCCGTTGACGGTGGAT	757						

RESULT 9

US-09-954-314-15

; Sequence 15, Application US/09954314

; Patent No. US20020127666A1

; GENERAL INFORMATION:

; APPLICANT: Rouviere, Pierre E.

; APPLICANT: Brzostowicz, Patricia C.

; TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIA

; FILE REFERENCE: BC1001 US NA

; CURRENT APPLICATION NUMBER: US/09/954,314

; CURRENT FILING DATE: 2001-09-17

; PRIOR APPLICATION NUMBER: 60/120,702

; PRIOR FILING DATE: 1999-February-19

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 15

; LENGTH: 10629

; TYPE: DNA

; ORGANISM: Brevibacterium sp HCU

US-09-954-314-15

Query Match 8.6%; Score 70.2; DB 10; Length 10629;

Best Local Similarity 46.2%; Pred. No. 4.6e-10;


```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9262
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(732)
US-09-815-242-9262

Query Match
Best Local Similarity 7.4%; Score 60.6; DB 10; Length 732;
Matches 278; Conservative 0; Mismatches 314; Indels 9; Gaps 2;

QY 31 CTAGAAGGAAAGTTCGCCCTTATAACAGGAGGAGCCAGTGGAAATTCGAGAAACACAGCA 90
Db 7 CTAGAACAATAAATATCTTTATACAGGTCGAGTCGGGAATGGTCTGCCATCGCC 66

QY 91 AAACCTCTTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGCCAAGATGAATTAGGT 150
Db 67 CACAAGTTTGGTCAACAGGAGCCACATTTGCTTAAACAGTCGTGGGCAATCTCAGTA 126

QY 151 CACTCAGTTGTCGAG---GCCATTGGCACTTCCAAATTCACCTACATCCACTGTGATGT 207
Db 127 GAATTCCTCGCTGAGTTTCAAACTATGGTATCAAGTGGTTCCTCCCATTTCCAGGAGATGA 186

QY 208 ACTAATGAAGACGGTGTAAATAATGCCGTGGGCAACACAGTTTCAACCTATGGAACACTG 267
Db 187 TCAGATTTTTCAGACACCTAAGCGTATGATGATCAAGCTATTGCAGAACTGGGTTTCAGTA 246

QY 268 GACATTATGTCAGCAATGCAGGAATTTCTGATCCCAACAGGCCCGCATCATAGACAAC 327
Db 247 GATGTTTTGGTCAACAAATGCAGGGAT-----CACCCAAAGATACTCTTTATGCTTAAGATG 300

QY 328 GAAAAAGCAGACTTTTGAACCGGTTCTCAGTGTAAATGTAACCGGAGTTTTTCCTATGCATG 387
```

```

QY 628 G 628
Db 601 G 601

RESULT 12
US-09-815-242-9483
; Sequence 9483, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9483
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(732)
US-09-815-242-9483

Query Match
Best Local Similarity 7.4%; Score 60.6; DB 10; Length 732;
Matches 278; Conservative 0; Mismatches 314; Indels 9; Gaps 2;

QY 31 CTAGAAGGAAAGTTCGCCCTTATAACAGGAGGAGCCAGTGGAAATTCGAGAAACACAGCA 90
Db 7 CTAGAACAATAAATATCTTTATACAGGTCGAGTCGGGAATGGTCTTGCCATCGCC 66

QY 91 AAACCTCTTCCCAACATGGAGCCAAAGTTGCCATTGCTGATGCCAAGATGAATTAGGT 150
Db 67 CACAAGTTTGGTCAACAGGAGCCACATTTGCTTAAACAGTCGTGGGCAATCTCAGAA 126

QY 151 CACTCAGTTGTCGAG---GCCATTGGCACTTCCAAATTCACCTACATCCACTGTGATGT 207
Db 127 GAATTCCTCGCTGAGTTTCAAACTATGGTATCAAGTGGTTCCTCCCATTTCCAGGAGATGA 186

QY 208 ACTAATGAAGACGGTGTAAATAATGCCGTGGGCAACACAGTTTCAACCTATGGAACACTG 267
Db 187 TCAGATTTTTCAGACACCTAAGCGTATGATGATCAAGCTATTGCAGAACTGGGTTTCAGTA 246

QY 268 GACATTATGTCAGCAATGCAGGAATTTCTGATCCCAACAGGCCCGCATCATAGACAAC 327
Db 247 GATGTTTTGGTCAACAAATGCAGGGAT-----CACCCAAAGATACTCTTTATGCTTAAGATG 300

QY 328 GAAAAAGCAGACTTTTGAACCGGTTCTCAGTGTAAATGTAACCGGAGTTTTTCCTATGCATG 387
```



```

; LENGTH: 805
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-797

Query Match          6.9%; Score 56.8; DB 10; Length 805;
Best Local Similarity 47.2%; Pred. No. 9.5e-07;
Matches 252; Conservative 0; Mismatches 267; Indels 15; Gaps 2;

QY 269 ACATATGTTTCAGCAATGCGAGGAATTCATCCCAACAGGCCCGCATAGACAACG 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 ACATCTAGTTAACTACGTGGAATCTCAATCCCAATTAACCTACCTACGCGCAACACTC 343

QY 329 AAAAAAGACAGCTTTGACCGGTTCTCAGTGTAAATGAATGACCGGATTTTCCATGCA 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 CCATTGAAGAATTCGATCGCATCTTCAAGGTGAACAACAGAGGATCATCTTATGCTGA 403

QY 389 AGCACGACGACGCTGTATGATTCACGACGAGCGAGTGGCAACATAATTTCCACTGCTAG 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 AAGAACGACAAAAGGCTAAACGTTGAGCGGCTGGTAGGATTAATCTACTGCTAACGCTCG 463

QY 449 TAAGCTCAACTATGCTGTGTTCTTACATGCCTATTGTTGTTCAAAAGCATGCTGTGT 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CGTTACCGAGGGTTAAATCCCGGCGCAGGAGCTTATACAGCATCAAGGCGAGCTGTG 523

QY 509 TAGCCCTTACTAGGAATCTGGCAGTCTGAGTCTGGACAATTTGSCATTAAGGTTAATTGT 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 AAGCAATGTTGAAGATCTTGTCTAAGGAATTAAGGTTTAGGCATCACTGCAAACTGTG 583

QY 569 TGTCTCTCTTCGGGCTCTCTACGGCTTTAGGCAAGAAATTTTCAGGGATTAATAATGAAG 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 584 TATCTCCAGGCGCTGTGTGGACGG-----AGATGTTTTTTGACGGGAAGAGTGAAG 634

QY 629 AAGAAATTTGAGAAATGAATAAACTTTTCGGGAAATTTGAAAGTCCAAAATTTAATGTTG 688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 AGACGCTGATGAATCA-----TTGAGAGGAGTCCCTTTGGTAGGCTTGGTGAGACTA 688

QY 689 AGGATTTGCCAATGCACTCTTTATCTGCTAGTAGTAGTGAGGCAAAATACGTGAGTGGAC 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 689 AAGATATTCTCTCTGTTGTTGTTCTTAGCTAGTAGTGTGAGTGTGATGATCAATGGAC 748

QY 749 ACATCTGTTTCATGATGAGGTTTCAGGCTCGCAATTCGTGAATCAAAAGTGT 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 749 AAGTATTGTTGTAATGCTGATTCCTCAAAATGAATATGAGTGTATGTTGTTT 802
```

```

RESULT 15
US-09-974-300-2259
; Sequence 2259, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2259
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2259
```

```

Query Match          6.7%; Score 54.8; DB 10; Length 762;
Best Local Similarity 48.4%; Pred. No. 3.5e-06;
Matches 186; Conservative 0; Mismatches 192; Indels 6; Gaps 1;
```

Search completed: February 18, 2003, 18:33:02
Job time : 131 secs

THIS PAGE BLANK (USPTO)
THIS PAGE BLANK (USPTO)